Task 1: Data Analysis and Simple Visualization

Task Overview

What you'll learn

- The importance of gaining meaningful insights from data and communicating your insights in a clear and concise way
- Demonstrate basic data analysis and visualization techniques in Excel using pivot tables

What you'll do

- Analyze a set of data about hours and costs for a set of projects
- Create a PowerPoint presentation to communicate your findings

---

Here is the background information on your task

One of the most frequently used and most important skills across the many roles at JPMorgan Chase & Co. is data analytics – the ability to gain meaningful insight from large volumes of often messy data and communicate your insights in a clear and concise way.

Over the years, many tools have been developed to assist with every stage of the data analytics lifecycle. Twenty years ago, the only tool commonly available was Excel. Today there is a wide and robust spectrum of tools used across every stage of the analytics lifecycle. Most analytics processes boil down to some version of this lifecycle:

- **Preparation** - Collect and aggregate data from various sources.
- **Exploration** - Understand the data and its characteristics and limitations.
- **Transformation, also called "munging"** - Clean the data, transform it to more useful forms, add calculated variables and statistics. Manage missing data and outlier data.
- **Model** - Build predictive / forecasting models, diagnostic models, and other mathematical analyses of the data.
- **Validation** - Test the model to validate it. Improve the model as needed.
- **Reporting and Visualization** - Make the data and the models that explain or use the data for prediction easy to see and understand by others. Create dashboards and other visual tools.

In this task you will do some simple data analysis and visualization using Excel, which is still one of the most commonly used analytics tools.

One of the most useful functions in Excel for robust data analysis is the pivot table. A pivot table in Excel is simply an automatically generated table of values that summarize aggregated data from another table in Excel. These summaries typically calculate sums, averages, or other statistics about the aggregated data.

If you have worked with pivot tables in Excel before, this task will be fairly straightforward for you. If this is your first time learning about pivot tables, now is the time to learn the basics. Use the links in Additional Resources, below, to go through some quick online tutorials about pivot tables. In about 15 minutes, you will learn the basic concepts and how to implement them in Excel.

Once you have done that, take a look at the example spreadsheet in Additional Resources, below, which demonstrates some simple pivot tables related to sales transaction data for a Pizza restaurant. The first tab holds the raw data, with each row representing a specific sale of a pizza, including the pizza type, who sold it, the sale price, and the day it was sold. The remaining tabs each use a pivot table designed to gain insight into this restaurant's sales. Take a look at the pivot table examples to see how they can be used to answer these specific questions:

- Which type of pizza has the most total sales revenue? (Answer: Cheese)
- Which server had the most sales revenue in total? (Answer: Monica)
- Which day had the most sales revenue? (Answer: Thursday)
- Which server had the most sales revenue on Saturday? (Answer: Tom)
- Which type of pizza had the most sales revenue on Saturday? (Answer: Pepperoni)

As you can see, pivot tables are a great tool for gaining insights from a dataset. Feel free to experiment on this example spreadsheet to get a feel for how pivot tables are set up and utilized. You will use this same technique when completing your task below.

Once you have done your analysis, it is time to communicate your insights visually. If you are already an Excel expert, the simple visualizations in your task below will be easy. However, if you are not already familiar and comfortable with graphing data in Excel and making charts to share with others, now is the time to practice. To help you get started, there are some links in the Additional Resources section below that will give you a quick introduction to graphing data in Excel. You can also find many other free resources on the web, including YouTube, if you would like to watch the process being done.

Your ability to successfully lead a discussion about the data and convince your audience of your recommendations depends on the clarity and impact of your communication. Your presentation should make it easy for your audience to understand the most important aspects of the data. Often, you will be the only one who has spent the time needed to be fully immersed in the data; your audience is depending on you to extract the most important data and findings for them. So make your presentation simple, concise, and direct. Tell your audience what the data set is, what it says, and what it means for whatever operational or business decision must be made relative to it. You want your audience to have confidence in your analysis and your conclusions, so be transparent about any problems, quality issues, or errors in the data, and how those may have affected your analysis. Then the discussion will naturally focus on what decision to make supported by the data, rather than on your analysis of it.

Here is your task

One of the many activities that JPMorgan Chase & Co. analysts support is the analysis of budget to actuals for things like projects or products. For this task, you will analyze a set of data about hours and costs for a set of projects, each with various tasks, resources, and resource rates. Your task is to analyze the data contained in the Excel spreadsheet in Additional Resources, below, and then write a short PowerPoint presentation to communicate your observations and key findings about that data. In your presentation, be sure to use the appropriate simple visualizations of the data, using the template provided in the Additional Resources below.

First, open the spreadsheet and familiarize yourself with the data. What kind of data is there? What information do the columns contain? What kind of trends could you see if you graphed that data in various ways? What kinds of observations are relevant to understanding budget to actuals? You could consider such questions as:

- Which projects are over or under budgeted hours?
- Which projects are over or under budgeted costs?
- Which personnel are over or under budgeted hours?
- Which personnel cost the most? The least?

But you don't need to limit yourself to those. You could analyze the dataset on other variables or trends, monthly cost variance, or which personnel is better at estimating their hours. Also, how could you analyze the data to see if there are any personnel who are charging an unrealistic number of hours in total? Does that tell you anything? Then, begin your formal analysis of the data. For this task, you are going to analyze the data using pivot tables for the questions above and any others you think are relevant or interesting about this data. You may need to filter, sort, or modify the spreadsheet so that you can see data subsets for different criteria and set up your pivot tables. You may need to use formulas to compute new variables to pivot on, such as defined buckets or sums or differences for some data.

You may also consider other ways to analyze the data. As you chart the data, look for trends, outliers, patterns, and other potential insights. What do your charts tell you about these projects and their performance?

Once you have analyzed the data, use the PowerPoint presentation template to communicate your findings. Add in your chosen charts from Excel, and then write your observations and key findings about the data in bullet point format, summarized on the last slide. Feel free to modify the template and add or remove slides as needed to fit your chosen data analysis. But remember, your job is to make the data easily understandable, with observations communicated directly and concisely. Make sure you title your charts with specific descriptions of the data in the chart and title the slides with a summary of what the chart shows. And keep your presentation short, with no more than 5-7 slides for presenting your data analysis.