Excel Data Science Projects

Tech Talent South (original author)
Ai Yukino (PDF rewrite)

Excel Project

Let's put some of what you learned together into three projects!

You may want to split up and work in groups, as today's projects are easier when worked on together.

We will be performing some basic data analysis in Excel on some clean datasets and then submitting our projects for grading. Let's get started!

Excel Project

Our projects are adaptations of projects from , however our goal for these projects should be to work as a team in order to demonstrate as many of our collective skills with Data in Excel as possible, while learning new skills when applicable.

If you have any questions throughout today's project, you can visit the resource above, or ask your instructor or TA for assistance.

Project 1: Case Scenario

Often times we perform Data Analysis for a variety of reasons, and it can be helpful to have a case scenario in mind as an example of a motivation to perform Data Analysis, and provide insight on your analysis. Consider this possible scenario:

One day at a family dinner, your mom and your uncle got into a debate about which basketball team is better. Your mom is a loyal fan of the State Team while your uncle is a fan of that Other Team, a big rival of State Team. Your cousin who plays a ton of basketball videogames chimes in that the Other OTHER State basketball team is the best.

Project 1: Case Scenario

As your family debates which team is better, you listen and note the insights of the people around you, as you know it may help you to increase your understanding of your data.

You mention that there are multiple different ways you can compare the performance of the basketball teams and you will bring in your analysis to the next family dinner. You thank everyone for their contributions to the discussion, and ask them if they think you should pay attention to anything in the data in particular.

Gathering Data

You go home, and Download the following CSV (comma separated value file) from Kaggle, a wonderful resource for open-source data: https://www.kaggle.com/ionaskel/nba-games-stats-from-2014-to-2018

You spend some time looking at the data, and make some basic initial observations. This contains data from NBA games in the 2014-2018 seasons.

Your task for your first project is to:

Select three comparable NBA teams (to represent the three NBA teams from the earlier scenario) to look at.

Data

For each of the three teams, locate statistics that represent each team's performance.

At the minimum, please include the following performance indicators for each team:

1. Number of games won
2. Number of games lost
3. PTS – Total Points
4. G – Games (i.e., number of games played during the season)
5. FG% – Field Goal Percentage
6. 2P% – 2-Point Field Goal Percentage
7. TOV – Total Turnovers

Data

Please write an Excel formula to calculate the following statistics for each team:

1. Win-lose percentage (W-L%) = Number of games won / Total number of games played

2. Average points per game = Total Points / Total number of games played

3. Average Turnovers per Game = Total Turnovers / Total number of games played

Data

Provide TWO additional performance indicators that you think are appropriate.

Use clear titles, labels, sheet name, and other appropriate spreadsheet formats (e.g., font type, size, colors, borders, etc.) so users can easily locate the information.

Include appropriate information on the Page Header and Page Footer.

Apply appropriate settings so that all information will be printed on one page.

Analysis

Here are some questions to consider for your analysis:

According to the data, which team had the highest Win-lose percentage (WL%)?

According to the data, which team had the highest Field Goal Percentage (FG%)?

What is the Average Points per Game for the State University, i.e., your Mom's favorite team, during the most recent season?

What is the Average Turnovers per game for the Other Team, i.e., your Uncle's favorite team, during the most recent season?

Analysis

According to the data, which team had the best performance in the most recent season when it comes to Turnovers? Note: In basketball, a turnover occurs when a team loses possession of the ball to the opposing team before a player takes a shot at his team's basket.

What were the two additional performance indicators that you selected?

Why did you choose these indicators and which team performed the best according to these indicators?

Analysis

According to your spreadsheet, which performance indicator do you think is the most important for this analysis and why?

Among the three teams, which team do you think performed the best during the season and why? Please explain your answer using information from the analyses you conducted on the spreadsheet.

Insight

Utilizing what visuals and data you gained from your exploration and analysis, write a clear description of your data exploration, model, analysis, and objective results, and how you formed your opinion. Be sure to write clearly and keep things easily understood by everyone, not just those with an understanding of math/statistics.

When you are done, you can submit a zip-file of your data, exploration, analysis, and insights for all of your projects from these slide decks.

Data Visualization

Great job! Now it's time for our next project! This one will be
focused on data visualization. Let's get started!

Data Visualization

Before we start, let's review chart types in Excel:

How to Select Best Excel Charts for Your Data Analysis Reporting

Choosing the Right Chart Type for Your Data

Case Scenario

National Park Services (NPS) is a federal agency of the United States government responsible for managing all U.S. national parks, many American national monuments, and many other conservation and historical properties.

Information about visitation at units of the NPS is publicly available and may be downloaded at https://irma.nps.gov/Stats/.

Case Scenario

You work for a company that specializes in guided tours of national parks across the United States. As an assistant to the Director of Marketing, you were asked to analyze some visit data at selected national parks.

Your boss is specifically interested in the number of visitors as well as the trend of recreational visits over the years. The information will be used to develop marketing strategies, promotional packages, as well as advertising campaigns for the next five years.

Project Requirements

Please prepare a data file for the analysis. The data are publicly available and can be downloaded from the NPS Stats website located at https://irma.nps.gov/Stats/.

1. From the NPS Stats Home page, navigate to the National Reports section and click the Recreation Visitation By State and by Park (1979 – Last Calendar Year) report. Select one or more states as indicated by your instructor and click View Report. Click the Export dropdown menu and select Excel. Save the exported Excel file.

2. From the NPS Stats Home page, navigate to the Park Reports section. Select a park (of your choice or as indicated by your instructor) from the dropdown list and click the Recreation Visits By Month (1979 - Current Calendar Year) report. Click the Export dropdown menu and select Excel. Move the worksheet containing the Park Report data so it is located on the same file as the State Report from step 1a and make sure to update the worksheet name accordingly. Repeat these steps as needed to pull data for other parks as indicated by your instructor.

Project Requirements

Using the data file prepared in the previous steps, please answer the following questions. Some questions might require you to create charts and/or graphs using Microsoft Excel.

1. What are the parks located in the state that you selected?

2. How many people visited the parks in the state that you selected during the most recent calendar year? Please create ONE chart that provides the answer(s) to this question.

3. What is the percentage of visitors at each location? In other words, what were the compositions of NPS visitors in the selected state by location? Please create ONE chart that provides the answer(s) to this question.

4. When was the most popular time, i.e., month, to visit the top three parks in the state during the current calendar year? Please create ONE chart that could provide the answer visually. This chart should show the number of visitors for each location for each month.

5. Discussion: What type of Excel chart did you choose to create in order to provide an answer to each question? Why did you select this chart type? Please explain your answer.

Project Requirements

Please keep in mind that, at the minimum, each chart should . . .

- ▶ Use an appropriate chart type to the data that you're trying to represent.

- ▶ Provide an accurate and complete answer(s) to the question.

- ▶ Include appropriate and meaningful chart elements (e.g., chart title, data labels, legend keys, etc.) without being too cluttered.

- ▶ Use clear and appropriate titles and labels without being too cluttered.

- ▶ Be placed on its own chart sheet with appropriate worksheet (i.e., tab) names

Data Analysis with Pivot Table

It's time for our final project! This time we'll be using pivot tables!

Case Scenario

Headquartered in Memphis, TN, Grenadier Super Store (GSS) specializes in office supplies and furniture. The company's customers range from individual consumers and small businesses (retail), to corporate organizations (wholesale) located in the United States and Canada.

You are an intern working for the Canada division of GSS. Your supervisor has given you an Excel file containing Order data from 2009-2012 and he would like you to analyze Orders/Customers/Sales data using PivotTables and PivotCharts.

Data

The data is included in an Excel file attached to this lesson.

Using the starting data file, please create PivotTables and PivotCharts that can be used to answer the following questions.

Analysis

What are the Regional Sales by Product Category and Product SubCategory?

Please create ONE PivotTable showing Total Sales breakdown by Region, Product Category, and Product Sub-Category.

Use information from the PivotTable to answer the following questions:

▶ What was the Total Sales figure included in this data set?

▶ Which Product Category had the highest sales?

▶ Which Region had the lowest sales? iv. What was the Total Sales of Appliances in Ontario?

Analysis

What are the Total Costs of Shipping by Order Priority and Ship Mode?

Please create ONE PivotTable showing the total Shipping Costs organized by Ship Mode and Order Priority.

On the same worksheet, please also create one a PivotChart (based on the PivotTable) to visually compare the shipping information.

Analysis

What are the Total Costs of Shipping by Order Priority and Ship Mode?

Use information from the PivotTable and PivotChart to answer the following questions:

1. What was the Total Shipping Cost for Critical orders?

2. GSS incurred the most shipping costs using which shipping method?

3. Discussion: If the Delivery Truck is the most economical but the slowest shipping method and Express Air is the fastest but the most expensive one, do you think the company appropriately spent shipping costs based on the Order Priority? Please explain your answer.

Analysis

Who are the most valuable customers?

Please create ONE PivotTable showing the Customer Names who placed orders with GSS during 2009- 2012.

For each customer, please also show the total number of orders, Total Sales, and Total Profit.

Add a Slicer or a Filter that can be used to show the information specifically for each Customer Segment.

Analysis

Who are the most valuable customers?

Use information from the PivotTable to answer the following
questions (Hint: Filter and sort the data in the PivotTable to locate
the answer):

1. Which Small Business customer had the highest sales?

2. Which Corporate customer placed the most number of orders
   in 2009- 2012? How many orders were placed by the Corporate
   customer?

3. Which Consumer customer was the most profitable one?

4. What is the sales figure of the least profitable Home Office
   customer?

Analysis

At the minimum, each PivotTable should...

1. Contain all required information

2. Be well structured in order to easily locate information and provide accurate and complete answer(s) to the question.

3. Use clear and meaningful headings and labels

4. Use an appropriate number format

5. Be placed on its own worksheet with an appropriate worksheet (i.e., tab) name.

6. The PivotChart should be on the same worksheet as the PivotTable. Make sure to use an appropriate chart type with complete and meaningful chart elements (e.g., chart title, data labels, legend keys, etc.) and clear and appropriate titles and labels without being too cluttered.