| **Milestone Project 1: Scenario 3**

Overview

You are the manager of a very successful bakery chain. Every few months, new items are introduced to the menu. In the past, the bakery chose to add items that could be made with whatever ingredients were leftover in the kitchen. However, you think this process could be improved upon. You recently got a hold of a dataset that shows sales records from the past few months at Keki's bakery.

Your report will provide information on the highest- and lowest-performing baked goods to persuade bakery stakeholders to adjust their inventory. Your report should also include information about the most profitable days for the bakery.

Audience: Bakery stakeholders

About the Dataset

- **Date:** Date of the sale
- **Day of Week:** Day of week of the sale
- **Total:** Total sales in South Korean Won, ₩
- **Place**: Name of bakery
- *The rest of the columns contain the items bought in each order.*

Part 1: Data Preparation

Tool: Microsoft Excel

Step 1: Define the goal.

What is the goal for this data analysis? What questions are you trying to answer?

The goal for this data analysis is to find the most profitable baked goods. We want to know the most profitable pastries and which times and days are most successful.

Step 2: Remove irrelevant columns.

Not every column will be useful for analysis. We can remove any columns where every row is the same. *Hint: An easy way to tell if every row is the same is to use the filter columns tool.*

What column(s) did you remove and why?

I chose to keep all the columns because each column can answer a question. I kept the time-stamps in case someone asks if there are more sales occurring in the morning or afternoon. This data can help us determine the optimal operating hours

for sales. I chose to keep the day of the week to determine which days are most profitable.

Even though there are 0 sales for the bakery items 'Mad Garlic' and 'Croque Monsieur', I chose to keep those columns. 0 sales is good data that can answer 'which items have the lowest sales? Which items can we remove from the bakery menu?'

Step 3: Identify typos.

Search for any typos that exist in the dataset in the *Day of Week* column and correct them.

I have found many typos and other errors. Spell check claims 'angbutter' is not a word. I do not know what "angbutter" is; a quick internet search reveals that it is a Korean dessert and a combination of butter and red bean paste, usually served on a starch such as a scone or croissant. *(Source: Daniel Oh at futuredish.com 'Seoul Bakery Best-Seller: Ang-Butter Scone!' 2/8/19. Accessed 4/8/22.)*

The software Spell-check says that "Thur" is wrong but it's an abbreviation that almost every English speaker understands, therefore I will change all the 'day of the week' data to the actual word instead of the abbreviation.

The information under "place" is in Korean. I do not know Korean but I do know that different stores will have a different arrangement of characters. I am going to leave the data because if someone asks which store is the most profitable, I want to be able to answer that question.

The cell containing "lemonade" has an unnecessary white space between 'lemon' and 'ade'. Vanilla is spelled wrong in the title "vanilla latte"; I changed it to have the proper two l's. "Berry ade" has an unnecessary white space. I changed it to be one continuous word. Meringue was spelled wrong and has been fixed. "Cheese cake" is usually spelled 'cheesecake,' therefore I have made it one word.

The column titled "Total" does not have any units. Is it USD? Or the South Korean Won? Based on context clues and row 2, six breakfast items do not usually total 24000.00 USD. If I use an exchange rate calculator, 24000 KRW Is equal to about 20.00 USD. This is plausible and allows me to conclude that the column 'Total' is in KRW. The symbol for KRW is ₩. I have added units to the title of this column to clarify.

I have changed the first row to 'proper' format because it looks cleaner and more professional.

I deleted the row with the number "1,293,000" under "total". This is an outlier because it is more than 40 times the other numbers and isn't reliable data.

With the outlier present, the average is 21173, the median is 18500 and the standard deviation is 26992. When the outlier is deleted, the average becomes 20647, the median is 18500 and the standard deviation is 7739

There are two errors in the time stamps. Most of the times occur between 11AM and 5:30PM. However, on the data of February 19th 2019, there is a time point of 10:13PM, and on October 27th 2019 there is a time point of 11:02 PM. If these were a mix up of AM vs PM, I can't know without double checking with the source. These rows were deleted because it's not reliable data.

With these inconsistencies deleted, the new Total Sales average is 20642.57 and the standard deviation is 7735.64

Step 4: Identify nulls and missing values.

Almost every dataset contains missing values. The goal is to handle such values uniformly throughout the dataset.

How did you choose to handle missing values and why?

It looks like we don't have data for 'place' from 7/11/2019 to 8/12/2019

It's possible we didn't start collecting that data until August 2019. Another possibility is that there was only one bakery in operation. And Keki's Bakery started expanding and opening new locations in August 2019. I chose to keep the data because it can still answer overall sales questions on which day and which items are most popular.

There is 'timedate' data in row 2655 containing '6/18/2020, 2:52:00 PM'. However there is no other data to go along with this datetime. Therefore, it's getting deleted.

Step 5: Remove duplicates.

Remove any duplicates in the dataset. Navigate to the **Data** tab in the Data Tools section and click the **Remove Duplicates** button.

Instead of deleting duplicate cells, I wanted to evaluate the data to ensure the duplicates are true errors or if they are reliable data. I selected only the first column for this task. I highlighted all the duplicate values using the 'conditional formatting' function. For the cells with duplicate date and time stamps, they take place at different locations, which is plausible. For the duplicate date/time values without 'place' data, there are different values for the 'total' data, which is also likely to occur. Therefore, I do not want to delete this data. If I simply use the 'remove duplicates' function, Excel will delete potentially important data. For example, when I execute the 'remove duplicates', my excel software says that 293 duplicate values were found and removed while 2358 unique rows remain.

Part 2: Data Exploration

Tool: Excel

Step 1: Use the SUM function in Excel.

We will define the most popular item as the one with the highest quantity sold. We can determine the total number of each item sold using the SUM function in Excel.

Your task is to determine the number of sales for each item as well as the most popular of the five.

Bakery Item Sales from 7/11/2019 to 5/2/2020

Bakery Item	Number Sold
Angbutter	3215
Croissant	1044
Plain Bread	1023
Tiramisu Croissant	944
Pain Au Chocolat	721
Orange Pound	565
Americano	513
Wiener	476
Pandoro	389
Cacao Deep	364
Jam	249
Vanilla Latte	241
Almond Croissant	230
Caffe Latte	214
Gateau Chocolat	210
Milk Tea	160
Cheesecake	92
Berryade	55

Meringue Cookies	49
Lemonade	38
Tiramisu	7
Croque Monsieur	0
Mad Garlic	0

The 5 most popular items are Angbutter, Croissant, Plain Bread, Tiramisu Croissant, and Pain au Chocolat.

The most popular bakery item is:

Angbutter

Step 2: Acquire column and row totals.

Use filters or a pivot table to get the total sales for each day of the week. This will tell us which day is the most profitable.

Sales from 7/11/2019 to 5/2/2020

Day of Week	Sum of Total Sales in South Korean Won, ₩
Sunday	11287500
Monday	6887000
Tuesday	56000
Wednesday	7574800
Thursday	8140700
Friday	6884700
Saturday	9098000
Grand Total	49928700

The most profitable day of the week is:

Sunday

Part 3: Gather Insights with Statistics

Tool: Excel Data Analysis ToolPak, Excel functions, and visualizations Step 1: Calculate and visualize descriptive statistics with the **Data Analysis ToolPak.**

Pick one month of bakery sales. Identify the most profitable and least profitable day in the month.

 I chose to analyze September 2019.

Date	Sum of Total Sales in South Korean Won, ₩
09/01/19	242200
09/02/19	132800
09/04/19	277200
09/05/19	247700
09/06/19	121700
09/07/19	285200
09/08/19	82300
09/09/19	215100
09/11/19	217700
09/12/19	185200
09/14/19	150700
09/15/19	293300
09/18/19	150800
09/19/19	301400

09/20/19	203000
09/21/19	301700
09/22/19	294200
09/23/19	174100
09/25/19	147400
09/26/19	54100
09/27/19	140400
09/28/19	287000
09/29/19	276700
09/30/19	113600
Grand Total	4895500

In September, the most profitable day was 09/21/19 where sales totaled ₩ 301700. The least profitable day was 09/26/19 where sales totaled ₩ 54100.
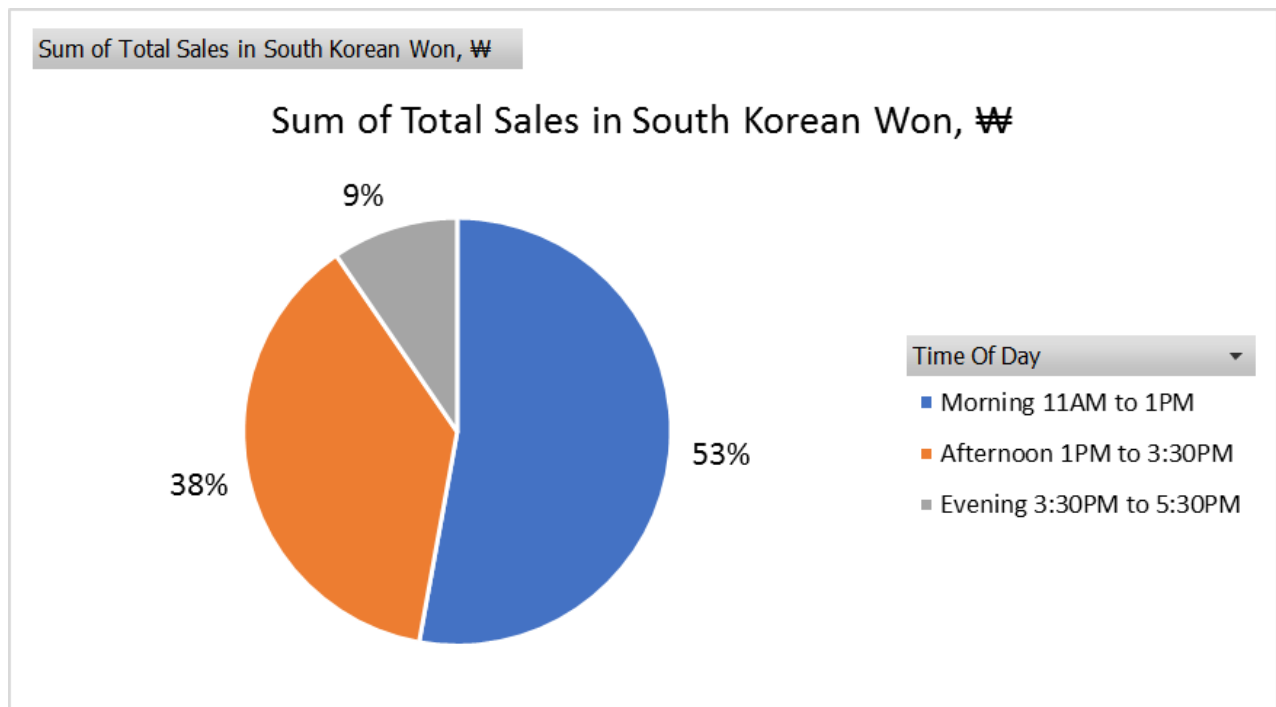
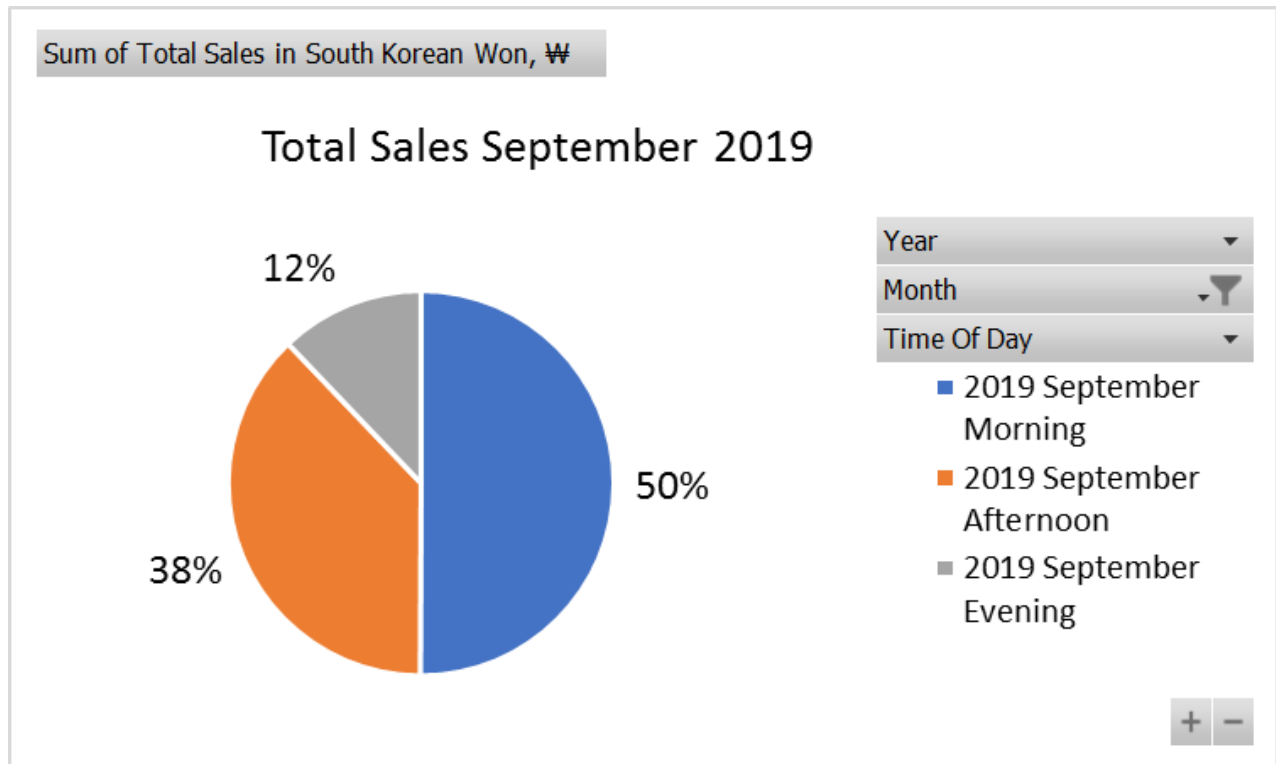What is the average amount sold for each day of the week?

September, 2019	Average of Total Sales in South Korean Won, ₩
Sunday	20854.39
Monday	19862.50
Wednesday	20871.05
Thursday	20747.37
Friday	21140.91
Saturday	21345.83
Grand Total	20831.91

The most profitable day in September 2019 was Sunday.

Create a breakdown of the sales for each timeframe: morning, afternoon, and evening. Since you only have a date-time column, you will need to create a new column to categorize the time of day. Place this data in a pie chart. Be sure to paste your visualization below.

In order to make this pie chart, I had to find the range of the time of sales. I discovered that the sales occur between 11:00 AM and 5:30 PM, based on the data. I then broke up the time into 3 categories. "Morning" sales occur between 11:00 AM and 12:59 PM. "Afternoon" is considered between 1:00 PM and 3:29 PM. "Evening" sales occur between 3:30 PM and 5:30 PM.

Sum of Total Sales in South Korean Won, ₩

Sum of Total Sales in South Korean Won, ₩

9%

38%

53%

Time Of Day

- Morning 11AM to 1PM
- Afternoon 1PM to 3:30PM
- Evening 3:30PM to 5:30PM

Sum of Total Sales in South Korean Won, ₩

Total Sales September 2019

12%

50%

38%

Year

Month

Time Of Day

- 2019 September Morning
- 2019 September Afternoon
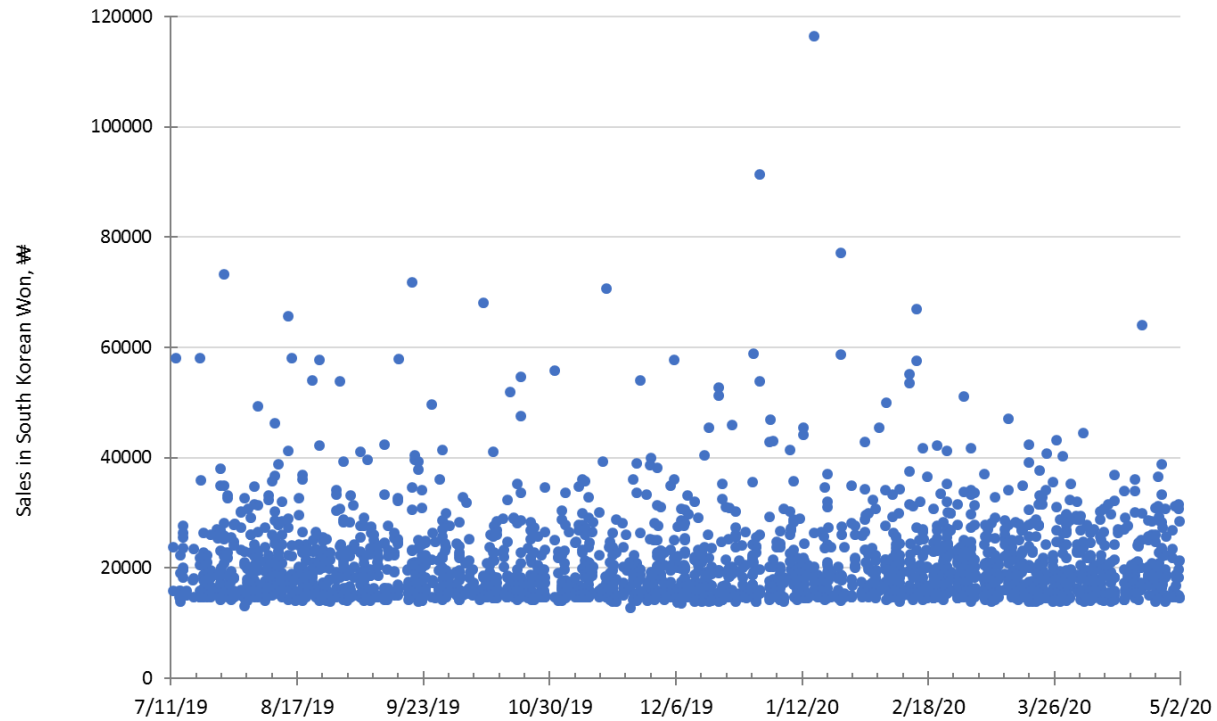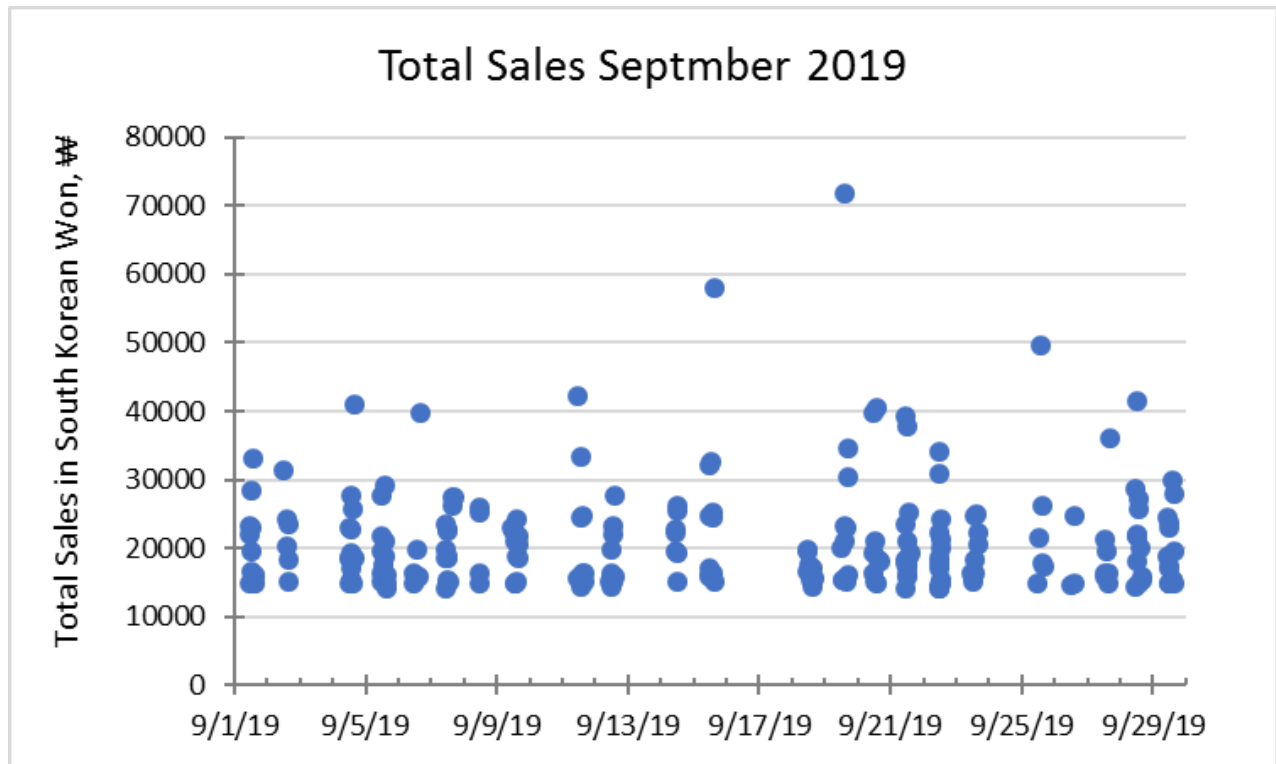- 2019 September Evening

+ −

Step 2: Create a scatter plot.

Create and interpret a scatterplot of *date* and *sales*. We want to show sales over time.

Be sure to paste your visualization below.

Total Sales From 7/11/19 to 5/2/2020

Total Sales Septmber 2019

Step 3: Use a combination of bar and line charts to compare groups.

Keki's Bakery is looking to maintain (and potentially increase) their high-traffic and high-sales days, as well as boost their sales on their lowest traffic days. We want to investigate which of the following pastries are doing the best and the worst:

- Angbutter
- Croissant
- Plain Bread
- Tiramisu Croissant
- Pain Au Chocolat
- Orange Pound
- Wiener
- Pandoro
- Cacao Deep
- Jam
- Almond Croissant
- Gateau Chocolat
- Cheese-Cake
- Meringue Cookies
- Tiramisu
- Croque Monsieur

- Mad Garlic

Create a bar chart of the "Highest Selling Baked Goods" for the entire dataset organized from largest to smallest. Report the top 3 performing baked goods, as well as the bottom 3. Be sure to paste your visualization below.

In this bar chart, we are looking at food items only. All the drinks have been removed so that we can focus on the baked goods.

Baked Goods Sold 7/11/19 to 5/2/20
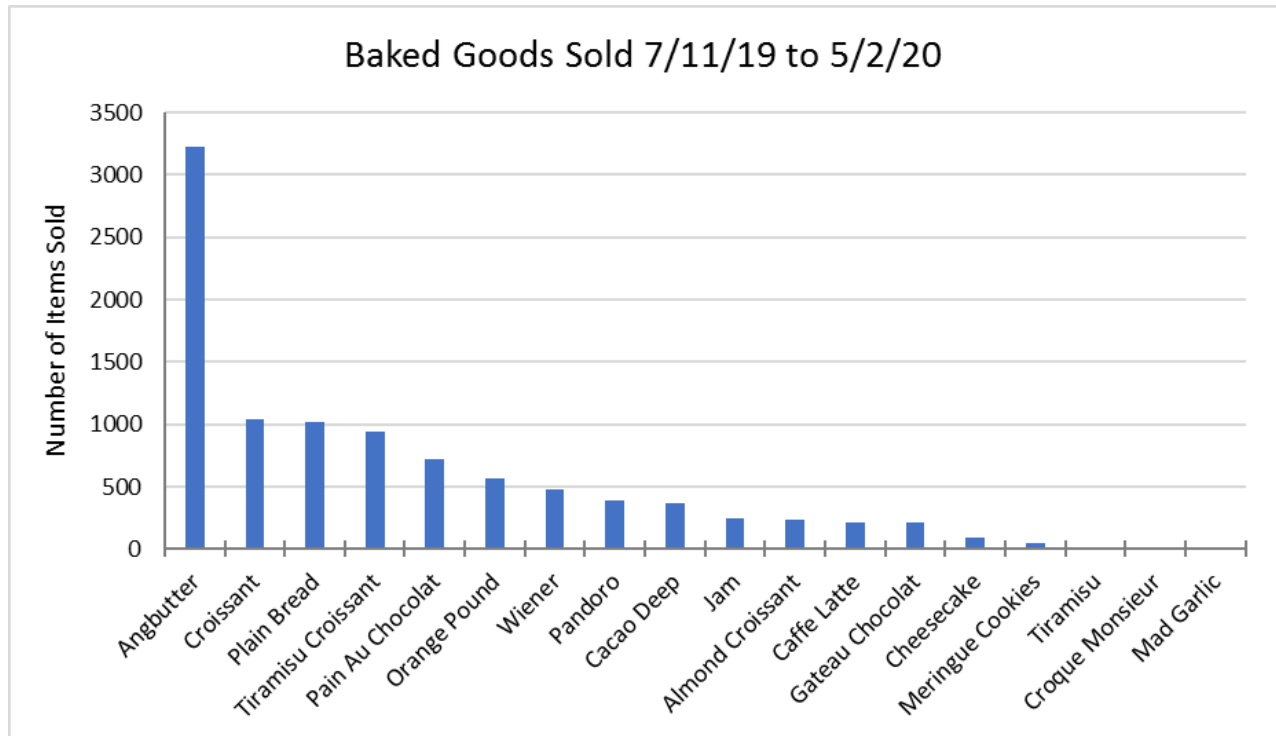
      The top 3 best-selling baked goods are angbutter, croissant, and plain bread.

      The 3 lowest performing baked goods, based on sales, are tiramisu, croque monsieur, and mad garlic.

Step 4: Make a business decision/recommendation based on the collected insights. Keki's Bakery relies on the following information to make business decisions:

- Day

- Number of pastries sold

- Type of pastry

What percentage of sales does the bottom ranking pastry make up for a given week?

Both Mad Garlic and Croque Monsieur make up 0% of sales during any given week. This is because these items have not been sold at all from 7/11/19 to 5/2/2020.

The entire data set begins on a Thursday so I broke up the weeks to begin from Thursday to Wednesday. During the week, starting from 8/8/19 to 8/14/19, two Tiramisu were sold. During the same week 330 total items were sold (including drinks). During this week Tiramisu consisted of 0.6% of all sales.

Open a new sheet in Excel and create a table for a month's worth (3 months) of one of the lowest pastry's sales.

Disclaimer: I have asked the teacher and my fellow classmates about this question. Everyone agrees this question is poorly worded and confusing. I have done my best to answer the question.

| Month	Sum of Meringue Cookies	Sum of Tiramisu
August 2019	9	2
September 2019	8	
October 2019	2	

**   Here is a table to help us better understand how the lowest 2 selling pastries perform month by month. (This chart does not include Croque Monsieur or Mad Garlic which have 0 sales) . Here we see that only 2 Tiramisu were sold and then declined to 0 for the rest of following months. Meringue cookies, which are the 4th least popular food item overall, did much better. However, it's still in the single digits and not very popular overall.**

Create a multi-line graph to track the pastry's performance throughout the month. Each line should be a day of the week. Be sure to paste the visualization below.

Disclaimer: I have asked the teacher and my fellow classmates about this question. Everyone agrees this question is poorly worded and confusing. Instead, I will use it as a guide to create a business insight and elaborate on why I chose this graph.
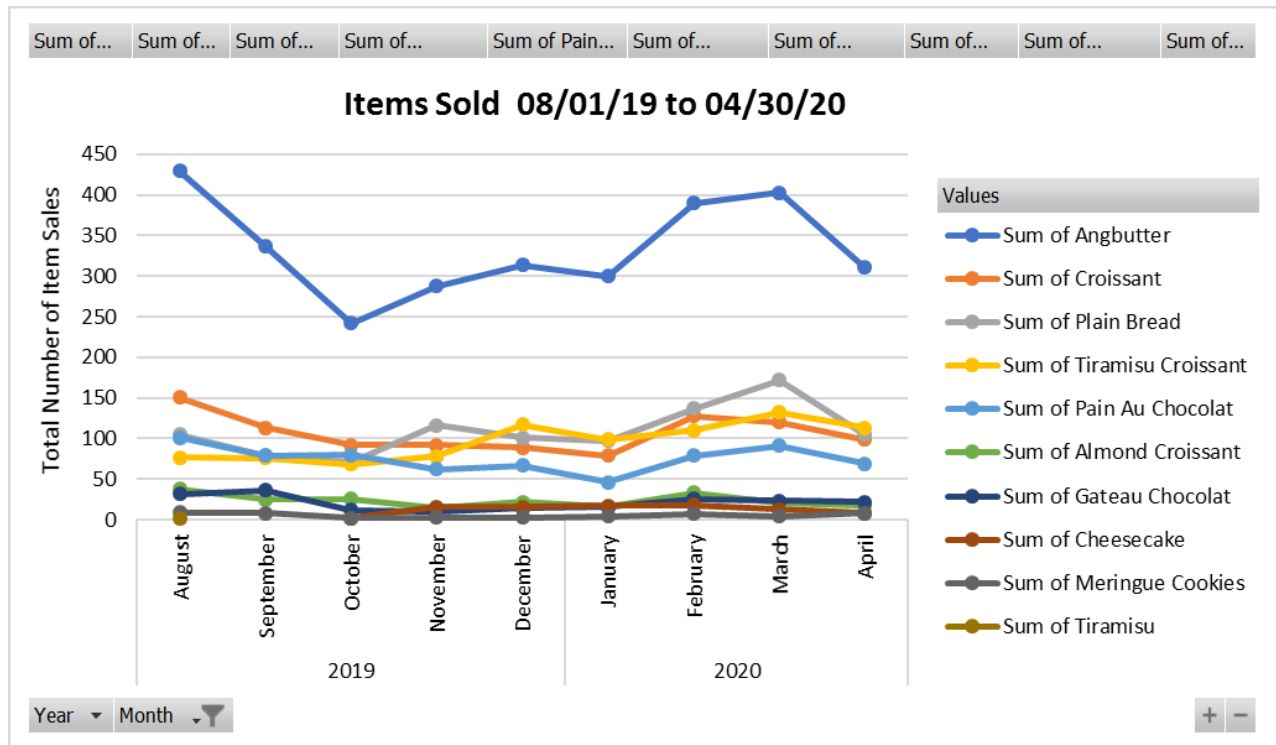
This graph summarizes the total sales by month of the top 5 most popular and the 5 least popular baked goods/pastries. July and May were not included because they are not complete months and don't give a good representation of an entire month of sales. The graph can help us make business decisions because we can see how or if the amount of sales are correlated with the month, weather, or holidays. In this graph there is not a noticeable change of sales between months. The only one that stands out is Angbutter which went through a steep decline in September and October and steadily climbed back up towards the same amount of August sales. But, it then drops in April.

Items Sold 08/01/19 to 04/30/20
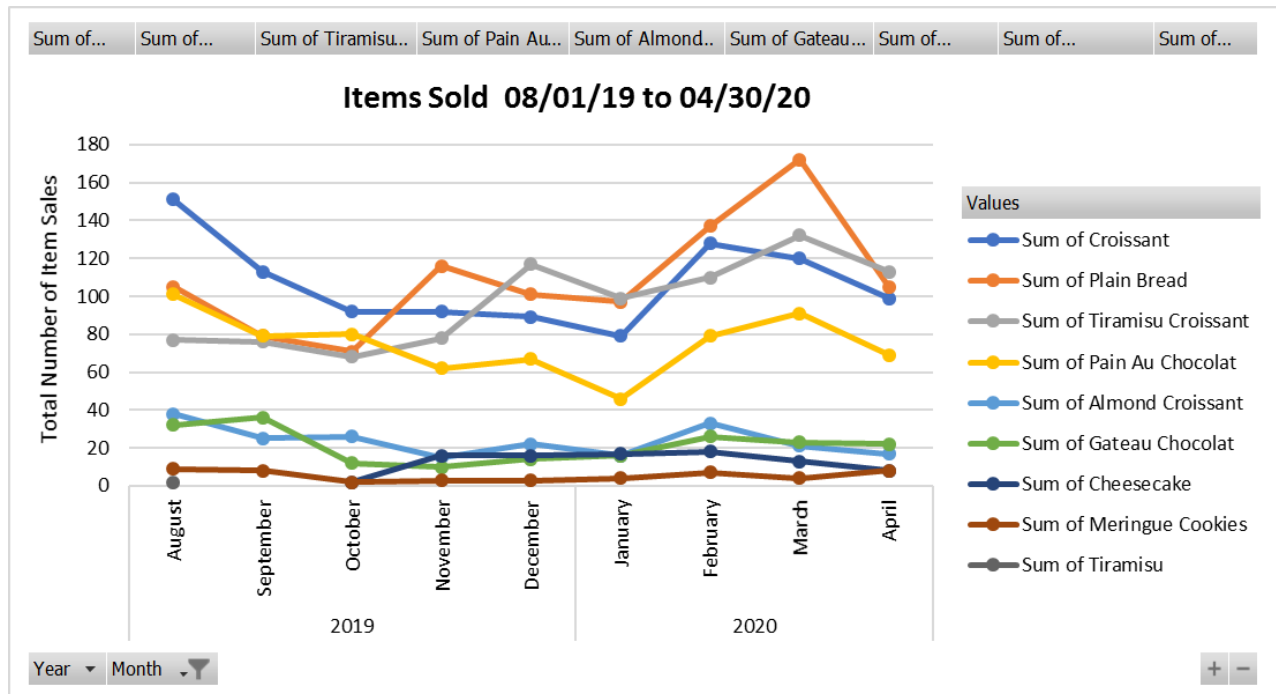
        If we remove Angbutter from the graph, it's much easier to see the change in the rest of the pastries because the x-axis now has a smaller range. In this graph, we can better see the other pastry sales in relation to one another. Croissant follows a similar pattern to Angbutter. Plain bread sales do not decline as steeply and then rise with a maximum in March, 2020.

        Only 2 Tiramisu were sold in August and then zero the rest of the following months.

Create a case as to whether Keki's Bakery should keep this pastry or get rid of it from its offerings.

        Keki's Bakery should get rid of the item Tiramisu from the menu. We do not sell enough to justify keeping it. And if we look at the cost of materials and labor compared to the revenue, we would most likely be losing money.

Part 4: Plan a Report

Tool: Word document, whiteboard application such as Miro
Step 1: Choose a report style.

Which report style will you use?

    A. Annual, quarterly, monthly

    B. Compliance

C. Progress

D. Feasibility

E. Operational

F. Strategic

G. Executive

H. Showcase a specific issue

I. Specific sector

Detail why you chose this option.

The report will be a mix of strategic and executive. This visual report is going to showcase our operations and predicted outcomes when we introduce new hours and change the menu.

Step 2: Gather report details.

Provide a title for your report based on the main goal or key insight.

Keki's Bakery: Menu and Hourly Adjustments

Write a brief description (about 2-3 sentences) on what your report is about.

This report highlights the need to remove items from the menu and change the hours. By making these small changes, Keki's should see a growth in sales.

Produce a list of everyone on the team and their roles such as "created visualizations" or "completed data preparation."

Sarah Patterson: Created visuals, completed data preparation and cleaning, wrote report

(This was an individual project, not a group project)

Step 3: Plan the visualizations.

What will be the main graphic or chart? It should be the most important insight you want to share.

The most important insight I want to share is the morning sales infographic. By showing this, I want stakeholders to be able to look at it and think about extending the morning hours to increase sales.

What will be the supporting graphics or charts? Keep in mind that you might need other visualizations to illustrate the main point and convince your audience.

Another visual I will include is an extrapolation of the total number of sales

(transactions) if we open the bakery at 10:00AM instead of 11:00AM. This is a 'prediction' of the data that's based on total sales from the entire cleaned database.

Are there any other topic-relevant images that you will add to the report for a visual boost? For instance, you might want to include an image of a pastry in your report.

I want to include a picture of an alarm clock and a pastry. The alarm clock will remind stakeholders that customers wake up earlier than 11am and a pastry will remind them that this meeting is about Keki's bakery (in case they have many other meetings to worry about).

Step 4: Report key insights.

List the main insights you found in your data.

Majority of all sales occur between 11:00 AM and 1:00 PM.

Angbutter is by far the most popular item: it makes up 30% of all sales

Tiramisu, mad garlic, and croque monsieur are the least popular.

What solution or conclusion will you be making? List the insights or data you gathered to support this.

Half of all sales (53%) occur between 11:00 AM and 1:00 PM. If we open an hour earlier we can alleviate the morning rush and make more sales.

Angbutter is popular but other items need promoting. Let's institute a buy one, get any other pastry half off. By having this special one day a week, we can get more foot traffic and sell more of the other items. We also need to remove the items that have no sales. They clutter up the menu and the ROI on these items puts us in the red.

Part 5: Develop a Data Story

Tool: Word document, whiteboard application such as Miro

Step 1: Complete the data story checklist.

What do you want to do with your dataset?

A. Inform/summarize findings of a study

The purpose of the data story is to inform the audience of my major findings which include: most of the sales occur in the morning, angbutter is the most popular pastry, and tiramisu, mad garlic, and croque monsieur should be removed from the menu.

B. Classify the data

Break the sales down by hour. Verbally inform the audience that the bar graphs represent sales by transactions. The pie chart represents sales in revenue.

C. Make a company decision or predict future results

Predict future results of opening the bakery 1 hour earlier with the option of closing the bakery 1 hour earlier.

D. Inspire/persuade people to act

Compare and contrast current total sales with predicted future sales. Make the argument that most of the overhead cost should remain about the same but have increasing profits.

Who is your audience?

My audience consists of the decision makers who are the executive managers of Keki's Bakery chain. My top employees/team leaders are the non-technical and discovery-minded type of audience who will also have access to the report.

Step 2: Organize your story points.

Choose some common story points for your data story. Write a few details on how you will illustrate these points.

● Change over time

● Relationship of two metrics

● Intersection (when one metric surpasses another)

● **Prediction**

I will predict the number of sales transactions between 10:00AM and 11:00AM with a 95% confidence interval.

● Compare and contrast

● **Drill down (general → specific)**

I will summarize the data from the number of sales transactions categorized into morning, afternoon, and evening. I will drill down the data and categorize by hour.

● Zoom out (specific → general)

● Cluster (values concentrated in an area)

● Outlier (data that lies outside the norm)

<u>Step 3: Create a story arc.</u>

What is the setting or context (i.e., who, what, where, when)? Include a hook or something to get the audience's attention.

This is the story I will use during my presentation to help drive my point:

(Here's a rhetorical question to start). How many of us are like our main character Leanne?

Our story starts at home with the alarm clock beeping at 9:00 AM to wake up Leanne, a late 20 year old woman who works as a bank teller. She needs to be at work by 11:00 AM because that's when the bank opens. She gets ready and leaves for work but needs to stop for breakfast. Leanne eats the same thing every day: an americano to get her caffeine fix and an angbutter on a croissant which has been her favorite pastry since she was 5. There is a Keki's Bakery located on the way from Leanne's apartment to her work. She wishes she could stop there because they make the best angbutters in all of South Korea but she can't because they don't open until 11:00 AM. She'll have to stop at Starbucks instead.

What are the rising insights that support/lead to your goal or main point?

Here we have lost a sale because most businesses open at 11:00AM or earlier. If we change our hours, we can get more foot traffic from commuters like Leanne, on their way to work instead of during their lunch hour.

(Here are rhetorical questions that I will ask during my presentation). How will we target customers like Leanne? What changes can we make to draw her into the store?

<u>Step 4: Add context to your story.</u>

Is there any background information the audience needs to know to make sense of the data insights?

Hours of Keki's Bakery which are currently 11:00 AM to 5:30 PM

Story Conclusion or hopeful outcome of my report.

After Keki's Bakery annual executive board meeting, they have decided to open all stores 1 hour earlier and if profits have increased by the end of the first quarter, they will open half of the stores 2 hours earlier. The board will reconvene after the second quarter to determine whether to open all stores at 9:00 AM or 10:00 AM, basing their decision on net profits.

Part 6: Build a Report

Tool: Excel

Create a one-page report (using the *Part 6_Report Template.xlsx*) that includes:

- Specific, targeted metrics illustrated with meaningful visualizations ●
  Storytelling techniques

- The recommendation or solution for the client

Consider the following when structuring your report:

- Report goal

Inform and persuade to change the menu and hours

- Color scheme

Blue orange white black

- Visualizations

Simple and easy to understand graphs. A picture of a croissant and an alarm clock to make it a little more fun and visually appealing.

- Text and graph balance

The final format must be an Excel document that your team will turn in, in addition to this packet.

Pictures are free domain. From the website https://publicdomainvectors.org/