Correlation Assignment: Excel Data Analysis Line Charts and Correlation Coefficients

David D Berberena

Bellevue University

DSC 500 Introduction to Data Science

Nasheb Ismaily

October 1, 2023

Correlation Assignment: Excel Data Analysis Line Charts and Correlation Coefficients

**Dataset 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Action Figure Sales | Clothing Sales | Food Sales |  |  |
| $265.00 | $1,286.00 | $49,728.00 |  |  |
| $498.00 | $1,193.00 | $53,639.00 |  |  |
| $687.00 | $1,008.00 | $68,236.00 |  |  |
| $346.00 | $1,429.00 | $44,971.00 |  |  |
| $502.00 | $998.00 | $55,643.00 |  |  |
|  |  |  |  |  |
|  | *Action Figure Sales* | *Clothing Sales* | *Food Sales* |  |
| Action Figure Sales | 1 |  |  |  |
| Clothing Sales | -0.793954521 | 1 |  |  |
| Food Sales | 0.907864518 | -0.848458635 | 1 |  |
|  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Interpretation of Dataset 1**

My thought process behind this dataset was a Japanese-style anime restaurant that sells action figures and anime-themed clothing. Based on the correlation coefficients generated in Microsoft Excel from the original dataset, clothing sales and action figure sales have a strong negative correlation. This means that while clothing sales decrease, action figure sales increase. Food sales and clothing sales also have a strong negative correlation, meaning that food sales take a hit while clothing sales spike. In this dataset, there is a strong positive correlation between food sales and action figure sales. With the coefficient being very close to 1, this means that food and action figure sales increase in the same direction.

With this and all correlation analyses, defining a correlation between two variables does not automatically imply causation. There are other variables outside the scope of this dataset that could be aiding in causing both food and action figure sales to increase together. There could be a special going on at the anime restaurant providing a nominal discount for those purchasing both food and an action figure within the same visit. A famous anime voice actor could be in town and customers want to buy apparel to have the voice actor sign when they go see him, driving up clothing sales. Buying and reselling unique action figures could be prevalent in the city where the anime restaurant is located, hindering clothing sales. The correlation coefficients show that any two combinations of these three variables are related in some unknown way, to some quantifiable magnitude, and for some unknown reason(s).

**Dataset 2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Diaper Sales | Formula Sales | WIC Allowance |  |  |
| $380.00 | $200.00 | $1,125.00 |  |  |
| $400.00 | $320.00 | $1,040.00 |  |  |
| $360.00 | $260.00 | $1,310.00 |  |  |
| $540.00 | $290.00 | $880.00 |  |  |
| $420.00 | $370.00 | $970.00 |  |  |
|  |  |  |  |  |
|  | *Diaper Sales* | *Formula Sales* | *WIC Allowance* |  |
| Diaper Sales | 1 |  |  |  |
| Formula Sales | 0.266010659 | 1 |  |  |
| WIC Allowance | -0.83678232 | -0.504372063 | 1 |  |
|  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Interpretation of Dataset 2**

The correlation coefficients generated from the second dataset are much more varied in their values than in the first dataset. This dataset was generated with the thought of new mothers caring for their infant children. The correlation between formula sales and diaper sales is a very weak positive one. This means that while they both increase together, the strength of their relationship is minimal to almost negligible. This could easily be attributed to the fact that many new mothers breastfeed their children unless specifically directed by their doctor. Cases like that do exist more commonly than what is initially thought, allowing for formula to be given to infants at an early stage. However, diapers are a necessity regardless of a woman’s ability to breastfeed their child, so the coefficient is properly justified in its weak correlation.

The correlation between a mother’s WIC allowance for any given amount of time and diaper sales shows a strong negative correlation, once again meaning that while one variable decreases (WIC allowance), the other increases (diaper sales). The relationship here does not necessarily suggest causality as it logically would make sense that if a woman’s WIC allowance is larger, they might buy more diapers to allow their child to have an ample supply. Even if the allowance were larger in that time period, a woman might already have enough diapers on hand and need other supplies for the child, implying that just the allowance has shrunk or grown for a given time does not automatically cause the mother to buy diapers proportionate to the allowance amount. The final coefficient defines the WIC allowance and formula sales as a moderately negative correlation. Outside variables not featured in this dataset could easily be affecting the relationship between these two topics, yet there is a noticeable relationship present. While they both increase in the same direction, other infant necessities could be muddling this correlation.

**Dataset 3**

|  |  |  |  |
| --- | --- | --- | --- |
| Total Number of Pokémon | Tournament Wins | Number of Shiny Pokémon |  |
| 697 | 3 | 18 |  |
| 34 | 7 | 0 |  |
| 108 | 4 | 2 |  |
| 336 | 2 | 5 |  |
| 1008 | 11 | 14 |  |
|  |  |  |  |
|  | *Total Number of Pokémon* | *Tournament Wins* | *Number of Shiny Pokémon* |
| Total Number of Pokémon | 1 |  |  |
| Tournament Wins | 0.456209042 | 1 |  |
| Number of Shiny Pokémon | 0.897030834 | 0.134945746 | 1 |
|  |  |  |  |
| |  | | --- | |  | |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Interpretation of Dataset 3**

My younger brother was the inspiration for this dataset. He has played Pokémon for as long as I can remember, and he became very interested in Shiny hunting and tournaments. The coefficients here are all positive, with varying degrees of strength. The first pairing is tournament wins and the total number of Pokémon, which carries a moderately positive correlation. That makes complete sense based on how the game is played in real life. Just because one has a plethora of Pokémon at their disposal to choose their tournament roster from does not necessarily mean they will win a lot of tournaments. There are many variables to consider with Pokémon games and tournaments, including, but not limited to, the use of hacked (overpowered) Pokémon, type advantage/disadvantage, move pool, level, stats, nature, and more. The number of Pokémon a player owns gives the player an unpredictability that may indeed help with winning tournaments, but the relationship has its outside variables that weaken the strength of the correlation.

The next two variables that correlate positively are the number of shiny Pokémon a player has and the total number of Pokémon they own. This positive relationship is very strong and is again justified by the nature of the games. A shiny Pokémon is one of a different color, and they are statistically rare to encounter. A player may go the entirety of the game without encountering one and will catch normal Pokémon along the way with the hope that they might find and catch a shiny one. It is very likely that a player could catch all catchable Pokémon in that particular game before obtaining even one shiny Pokémon. However, as the appearance of a shiny Pokémon is based on a percent value, another player could feasibly see and capture four or five of these rare creatures before finishing half the game! Most players do attain more Pokémon before encountering a shiny one, but more total Pokémon does not equal more shiny Pokémon, and vice-versa.

Finally, the number of shiny Pokémon a player has related to tournament wins is defined by the correlation coefficient as negligibly positive. The coefficient is so close to zero that there is almost no relationship between the two variables. As I’ve said before, the Pokémon competitive gaming scene justifies this correlation. Shiny Pokémon are not much different from regular Pokémon, albeit they are much more aesthetically pleasing (well, most of them anyway). They have similar strengths, weaknesses, and temperament. They do not have any more ability to aid in winning tournaments than any other Pokémon does. In order for this to even make sense logically through playing the games, you would have to catch a shiny Pokémon first (which statistically has been known to take a long time for most players), train them the right way to increase EV values, and max out their stats. Some players even go so far as to infect their Pokémon with a virus to make them stronger, shiny or not! While there is almost no correlation between these two variables, there is a small amount as can be seen by some players who quit the moment they see a rare Pokémon like a shiny one, or by those players who get away with legally altering their special creatures to win the tournament. Even with all that being said, the fact that one has a shiny or a whole team of them does not automatically cause a player’s tournament wins to increase.