Data Analysis in Spreadsheets

Predefined mathematical functions

- ROUND(number, [places]) = Round number to the number of decimal places, default is 0
- SQRT(number) = Square root of number
- MIN/MAX(data range) = Minimum/maximum value in a range of numbers
- RANK (value, data range, [is ascending]) = Where a value ranks within a data range
- AVERAGE(data range) = Mean value for a range of data
- MEDIAN(data range) = Median value for a range of data

=ROUND(123.456)	=SQRT(100)	=MIN(A2:B2)	=MAX(A2:B2)	=RANK(A2,A2:D2)	=AVERAGE(A2:E2)	=MEDIAN(A2:F2)
123	10	10	123	1	66.5	38.25

Predefined string functions

- LEFT/RIGHT(string, [number of characters]) = Get characters from left/right of string. If something other than string is given, convert to string
- LEN(string) = How many characters long is string. If something other than string is given, convert to string
- SEARCH(search value, to search) = index position of first occurence of search value in item to search
- CONCATENATE(string, [other strings]) = combines all strings into a single string

=CONCATENATE("This", " ", "is a ", "test")	=LEFT(A2, 4)	=RIGHT(A2, 4)	=LEN(A2)	=SEARCH(" ", A2)
This is a test	This	test	14	5

Predefined date-time functions

- WEEKDAY(date, [type]) = returns day of week as integer for a given date. Can specify how days of week are numbered with **type**
- DATEDIF(start date, end date, unit) = returns the time difference between the end and start date in the specified units
- NOW() = Current date and time

=NOW()	=WEEKDAY(A2)		=DATEDIF(C2, A2, "D")
2/18/2020 10:50:57	3	2/18/2019	365

Conditional functions

- IF(logical_expression, value_if_true, value_if_false) = returns one of two values depending on how the logical expression resolves. If multiple conditions need to be evaluated, can place IF within result value of another IF
- COUNTIF(range, criterion) = count values that fit criteria within a range
- SUMIF(range, criterion, [sum_range]) = sum values that fit criteria within a range, can select different column to sum from column to check
- AVERAGEIF(range, criterion, [average_range]) = same as SUMIF for mean

	=IF(A2="a", 1, 0)	=COUNTIF(A2:A6,"=b")	=SUMIF(A2:A6,"=1")	=AVERAGEIF(A2:A6,"=1")
a	1	2	3	1
a	1			
b	0			
b	0			
a	1			

Lookups

- VLOOKUP(search key, range, index, [is sorted]) = vertical lookup. Select values from index column for rows where first column contains search key. Index is the column number, starting with 1, from the left
- HLOOKUP(search key, range, index, [is sorted]) = horizontal lookup. Select values from index row
- XLOOKUP(search key, lookup array, return array) = similar to VLOOKUP, except specifies lookup and return arrays independently
- SUMPRODUCT(array1, [array2, …]) = if multiple arrays are passed, take the product of the elements across arrays, then the sum of products. If only one array is passed, just the sum of elements is returned.

1	5	=XLOOKUP(4,A1:A4,B1:B4)	=VLOOKUP(2,AO27:AP30,2)
2	6	8	6
3	7	=SUMPRODUCT(AO27:AO30,AP27:AP30)	=HLOOKUP(5, AO27:AP30,4)
4	8	70	8

Exercises

1. Write a formula to **RANK** each **BMI**, within the column. Did you get the result you expected? Why did you get this result?
2. Use multiple **IF** statements to check if the **height** and **weight** are **NA**. If either are **NA**, output **0**, otherwise output the **BMI**. **HINT:** your formula may look something like **=IF(___, ___, IF(___, ___, ___))**
3. Calculate the **average BMI** for the values that are not **0**.
4. The **age** column is in years. Add a new column that is **age in days**, called **age(days)**. Add another column called **birthdate** that subtracts **age(days)** from the current time.
5. Add a new column that concatenates the list of movies with a **“ ,”**. Then add another column called **first_movie** and write a formula that uses string functions to return just the first movie the character appeared in.
6. Use a lookup function to find the character closest in age to you.