Data Analysis in Spreadsheets

Predefined mathematical functions

- ROUND(number, places) = Round number to the number of decimal places
- 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

	A	B	C	D	E
1	=ROUND(123.456, 0)	=SQRT(100)	=MIN(A2:B2)	=MAX(A2:B2)	=RANK(A2, A2:D2)
2	123	10	10	123	1

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

	A	B	C	D	E
1	=CONCATENATE("This", " ", "is a ", "test")	=LEFT(A2, 4)	=RIGHT(A2, 4)	=LEN(A2)	=SEARCH(" ", A2)
2	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

	A	B	C	D
1	=NOW()	=WEEKDAY(A2)	(Just an older date)	=DATEDIF(C2, A2, "D")
2	2/25/2020 13:59:09	3	2/18/2019	372

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

	A	B	C	D	E
1		=IF(A2="a", 1, 0)	=COUNTIF(A2:A6,"=b")	=SUMIF(B2:B6,"=1")	=AVERAGEIF(B2:B6,"=1")
2	a	1	2	3	1
3	a	1			
4	b	0			
5	b	0			
6	a	1			

VLOOKUP

- VLOOKUP(search key, range, index, [is sorted]) = Vertical lookup. This will look for a value (search key) in the **first column** of a data range (range) and return the value in the same row in a **different column** (index). Index is the number of the column you want returned, **starting from the first column as 1**. Optionally, you can ask for the **closest match (TRUE)** or an **exact match (FALSE, default)** in the is sorted argument.

	A	B	C	D
1	1	5	a	=VLOOKUP(2, A1:C4, 2)
2	2	6	b	6
3	3	7	c	=VLOOKUP(2.75, A1:C4, 2, TRUE)
4	4	8	d	6

HLOOKUP

- HLOOKUP(search key, range, index, [is sorted]) = Horizontal lookup. Similar to VLOOKUP, except searching along the other axis. Look for the **search key** in the first row of the data **range**, then give back the value in the same column for the indicated row (index). **Index** is the row number, starting with the top row of the data range as 1. Is sorted again looks for the **closest match (TRUE)** or the **exact match (FALSE, default).**

	A	B	C	D
1	1	5	a	=HLOOKUP("a", A1:C4, 2)
2	2	6	b	b
3	3	7	c	=HLOOKUP(5.1, A1:C4, 3, TRUE)
4	4	8	d	7

XLOOKUP (EXCEL ONLY)

- XLOOKUP(search key, lookup array, return array) = Similar to VLOOKUP, except specifies the lookup column and the return column independently. **Search key** is the value you are looking for in the **lookup array**. The value returned is the value at the same row position in the **return array**. (https://support.office.com/en-us/article/xlookup-function-b7fd680e-6d10-43e6-84f9-88eae8bf5929)

	A	B	C	D
1	1	5	a	=XLOOKUP("b", C1:C4, A1:A4)
2	2	6	b	2
3	3	7	c	
4	4	8	d	

SUMPRODUCT

- SUMPRODUCT(array1, [array2, …]) = Can accept 1 or more arrays. If only 1 is passed, returns the sum of the elements in the array. If multiple are passed, multiplies elements at the same index position of each array, then sums the products.

	A	B	C	D
1	1	5	a	=SUMPRODUCT(A1:A4)
2	2	6	b	10
3	3	7	c	=SUMPRODUCT(A1:A4,B1:B4)
4	4	8	d	70

Exercises

Use the starwars data for the following:

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. Use a lookup function to find the character closest in age to you.
6. BONUS: Add a new column called **first_movie** that contains just the first movie listed for each character.