

The background is a light gray gradient. It features several realistic water droplets of various sizes, some with highlights and shadows, scattered across the frame. In the upper center, there is a faint, circular, textured pattern that resembles a ripple or a lens flare.

# BUSINESS ANALYTICS

# DATA CLEANING

- DATA CLEANING IS AN ART THAT IS VERY IMPORTANT TO ANY DATA PROJECT.
- DATASETS USUALLY ARE NOT EXACTLY WHAT YOU ARE LOOKING FOR.
- DUPLICATE DATA, EXTRA CHARACTERS, MISSING DATA, ERRORS.
- GOOD DATA ANALYST HAS SKILLS TO PREPARE THE DATA AND LIMIT THE ISSUES.
- OLD SYSTEMS INCLUDE SYSTEM CHARACTERS WITHIN THE EXTRACTS THAT YOU DON'T WANT IN EXCEL.

# DATA CLEANING

- CLEAN() FUNCTION REMOVES ANY NON-PRINTABLE SYMBOLS IN THE DATA (E.G. ¶ )
- CREATE A NEW COLUMN AND USE THE =CLEAN(CELL REFERENCE) TO CLEAN EACH CELL, YOU STILL HAVE THE ORIGINAL DATA.
- TRIM() FUNCTION WILL STRIP EXTRA SPACES FROM TEXT, LEAVING ONLY SINGLE SPACES BETWEEN WORDS AND NO SPACES AT THE START OR END OF THE TEXT.

# DATA CLEANING

- SUBSTITUTE() FUNCTION USED IN CONJUNCTION WITH TRIM CAN REPLACE A SYMBOL WITH SOMETHING ELSE APPROPRIATE.
- FOR EXAMPLE CONVERTING EUROPEAN , IN DECIMAL WITH A .
- =TRIM(SUBSTITUTE(A22,"","."))
- INSTEAD OF USING FIND REPLACE WE CAN USE THIS METHOD AND WE RETAIN THE FORMAT OF THE ORIGINAL DATA.

# DATA CLEANING

- SOMETIMES DATA THAT IS IMPORTED INTO EXCEL COMES WITH FLAGS SHOWING ERRORS (GREEN TRIANGLE IN TOP CORNER OF CELL).
- WHEN YOU CLICK ON THE CELL AN ERROR SIGN POPS UP.
- WHEN YOU CLICK ON THE ERROR YOU CAN SEE THE NUMBER IS ACTUALLY STORED AS TEXT.
- IF A CELL IS NOT STORED AS A NUMBER WE CANNOT DO SOME OF THE MOST COMMON FUNCTIONS IN THE DATA SUCH AS SUM AND AVERAGE.

# DATA CLEANING

- =VALUE() CONVERTS A TEXT STRING TO A NUMBER
- OR HIGHLIGHT THE COLUMN CONTENTS AND ON THE DATA TAB CHOOSE TEXT TO COLUMNS, IN THE WIZARD, KEEP THE FILE TYPE AS DELIMITED AND CLICK NEXT, UNCHECK THE TAB AND OTHER DELIMITERS AND CLICK NEXT, CHOOSE GENERAL, ENTER THE DESTINATION CELL RANGE AND CLICK FINISH.

# DATA CLEANING

- =TEXT() CAN BE USED TO COMBINE A SET OF VALUES INTO A USEFUL SENTENCE

=“TODAY’S DATE IS “&TEXT(TODAY(),”MM/DD/YY”)

TODAY’S DATE IS 09/10/17

# DATA CLEANING

- CONVERTING TEXT TO COLUMNS ALLOWS YOU TO SPLIT A CELL/COLUMN OF TEXT E.G. FIRST NAME LAST NAME INTO TWO SEPARATE COLUMNS.
- DATA RIBBON
- USE THE DATA PREVIEW BOX TO CHECK THE OUTPUT IS AS EXPECTED.
- BE SURE TO CHOOSE THE APPROPRIATE DESTINATION FOR THE DATA YOU WISH TO ALTER.

# DATA CLEANING

- MANIPULATING TEXT IS AN IMPORTANT PART OF DATA CLEANING.
- TEXT FUNCTIONS ALLOW YOU TO MODIFY, MANIPULATE AND SEARCH TEXT DATA IN ORDER TO INCREASE THE USE OF THE DATA IN EXCEL.
- =RIGHT() FUNCTION ALLOWS YOU TO EXTRACT A SUBSTRING FROM A TEXT STRING STARTING AT THE END OF THE STRING FOR A DEFINED NUMBER OF CHARACTERS.
- =RIGHT("HELLO",2) RETURNS LO
- THE RIGHT() FUNCTION WILL COUNT SPACES AS A CHARACTER

# DATA CLEANING

- `=LEFT()` WORKS IN THE SAME WAY BUT FROM THE START OF A STRING.
- THIS IS USEFUL WHERE WE RECEIVE DATA THAT IS SOMETIMES REPEATED OR CLEARLY MADE UP OF PARTS OF DATA.
- FOR EXAMPLE AN IBAN NUMBER CONTAINS BANK INFORMATION AS WELL AS ACCOUNT INFORMATION.

# DATA CLEANING

- SOMETIMES YOU NEED TO CUT PART OF A PIECE OF DATA AT A CERTAIN POINT THAT IS NOT A POSITION BUT A CHARACTER.

VTYQ-2216 IF WE WANTED TO STRIP OUT 2216 WE NEED TO :

- FIND THE POSITION OF THAT CHARACTER.
- =FIND() FUNCTION RETURNS THE POSITION AS A NUMBER OF ONE TEXT STRING INSIDE ANOTHER.
- GET THE LENGTH OF THE STRING. =LEN()
- IF WE KNOW THE LENGTH OF A STRING WE CALCULATE THE NUMBER OF CHARACTERS WE WANT TO REMOVE

# DATA CLEANING

- IF C2 HOLDS THE STRING VTYQ-2216
- =RIGHT(C2,LEN(C2)-FIND("-",C2)) 1)
- PRODUCES THE RESULT 2216

# DATA CLEANING

- `PROPER()`
- `UPPER()`
- `LOWER()`
- `CONCATENATE()` PUTS MULTIPLE STRINGS TOGETHER

# DATA VALIDATION

- YOU CAN RESTRICT DATA INPUT USING CRITERIA IN ORDER TO REDUCE DATA ERROR AND INCONSISTENCIES.
- DATA TAB ON THE EXCEL RIBBON HAS DATA VALIDATION TOOLS