Week 4: Verifying & Reporting

Verification:

A process to confirm that a data-cleaning effort was well executed and the resulting data is accurate & reliable.

Involves:- Rechecking the work (cleaning work)
- Doing some MANUAL cleanups
- Taking a moment to sit back & think about the original purpose of data.

" Verification lets you catch the mistakes before you begin analysis "

The other part of this process is "Reporting".

" Reporting can show your teammates that you are 100 percent transparent about your data cleaning "

Impresses
SK !

Change Log:

A file containing a chronology ordered list of modifications made in a project.

"*Don't rush! Be accurate.*

Take pauses. Be Accurate."

--

The **First Step** is to going back to the original dataset (undeam) and compare it with what you have now.

- Like in the unclean data, there were nulls... see are they now?

- Like you corrected some spelling. Are all spelling corrected now?

___

The **Second Step** is to take a **Big-Picture View** of the project.

o Taking a big picture view of your project lets lets you to confirm you're actully focusing on the business problem

& your deata is capable of solving that problem.

What to do for Big Picture?

1. Consider the business problem.

2. Goal of the project.

3. Consider if deata is capable.

Tip:
Ask your teammate about review of the deatu & get fresh feedback.

" Like testing on whether the car is safe before letting it to run on the road ".

When the error can't be fixed manually.
Create Pivot table

Use ~~count~~ COUNTA function which is
• ___ value counts of
Pending

Use CASE SELECT
• ___ *

CASE
WHEN col = " "
THEN " "
ELSE " "
END

from here the same week-4 is
continued in another book.

Ayush ∞ Sheih

week 4 is continued from course 4.

So, what is overall included in
Data verification ?
· CHECKLIST.

...

Make sure you identified the most common
problems and corrected them following:

· Source of errors: ...

· NULL Data: Did you search NULLs with CF
and filtering too ?

· Misspelled Word: Did you locate misspellings?

· Mistyped numbers: Check numbers by visulizing
and sorting

· Extra spaces & characters: ...

· Duplicates: ...

· Mismatched Datatype: ..

· Inconsistent Strings: Made sure cell strings
are consistent and
meaningful ?

- Inconsistent Date format :

- Misleading column names : Oh, I do it first!

- Truncated data ?

- Business Logic : Did you check that whether your data makes sense given your business logic?

- - -

Now, that we have corrected everything we need *R to address what we corrected.

Through - Documentation :

The process of tracking changes, additions, deletion and errors involved in your or data cleaning effort.

Because Documentation
lets us to see :

- ~~Allow~~ Recover data-cleaning errors
  (instead of scratching head to
  try to recall ...)

- Lets you inform others of changing

- Determine quality of data

Change Log 1
          A file containing a chronologically
ordered list of modifications made to the
project.

Sheets : Version History
SQL : Query History

Make a file - Write changes there.