Google data analytics professional course

Week - 1

Data integrity and analytics objectives

Why data integrity is important

Data integrity is the accuracy, completeness, consistency, and trustworthiness of data throughout its lifecycle.

Data can also be compromised through human error, viruses, malware, hacking, and system failures

More about data integrity and compliance

Scenario: calendar dates for a global company

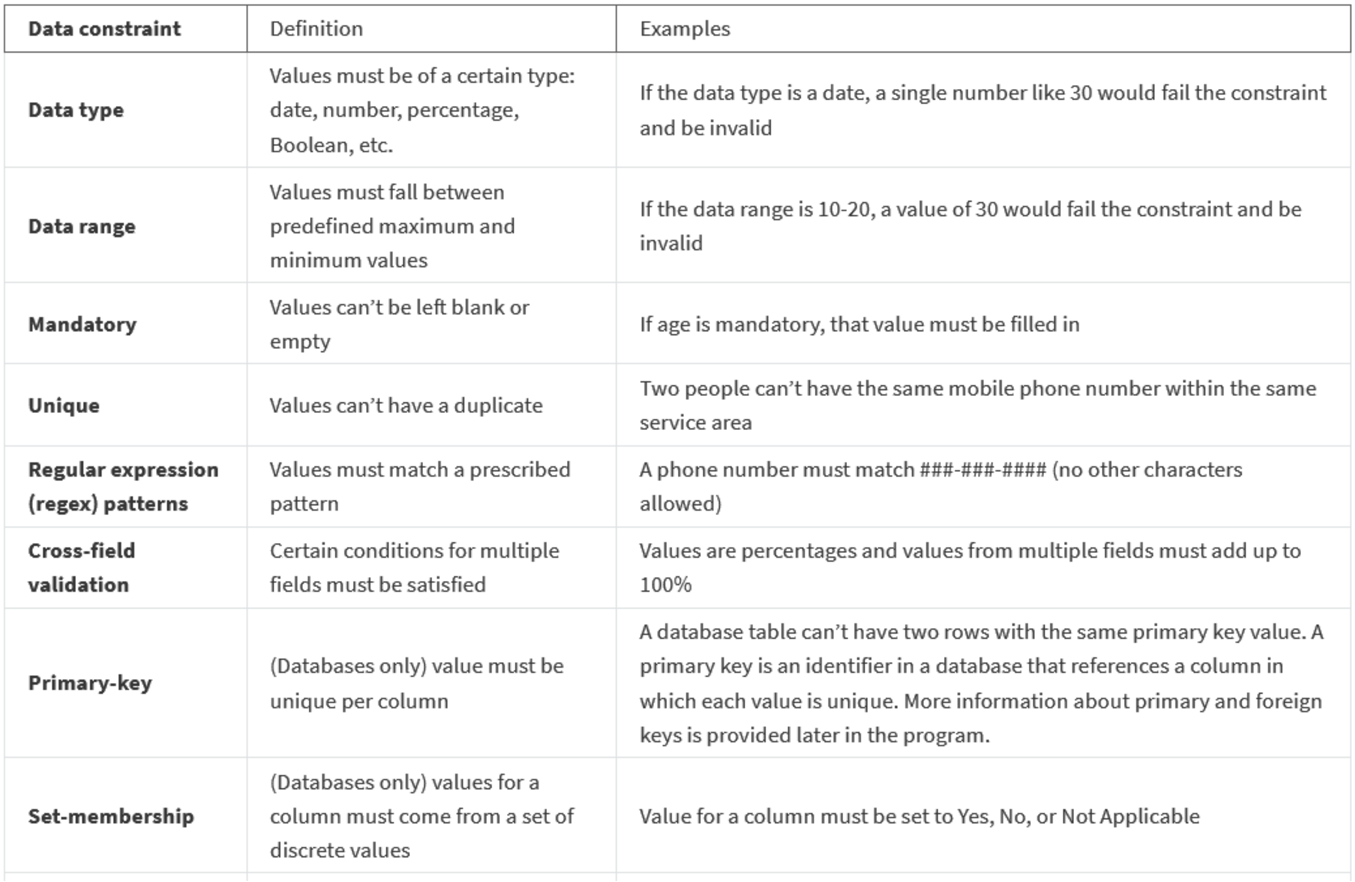
Calendar dates are represented in a lot of different short forms. Depending on where you live, a different format might be used.

- In some countries,**12/10/20** (DD/MM/YY) stands for October 12, 2020.
- In other countries, the national standard is YYYY-MM-DD so October 12, 2020 becomes **2020-10-12**.
- In the United States, (MM/DD/YY) is the accepted format so October 12, 2020 is going to be **10/12/20**.

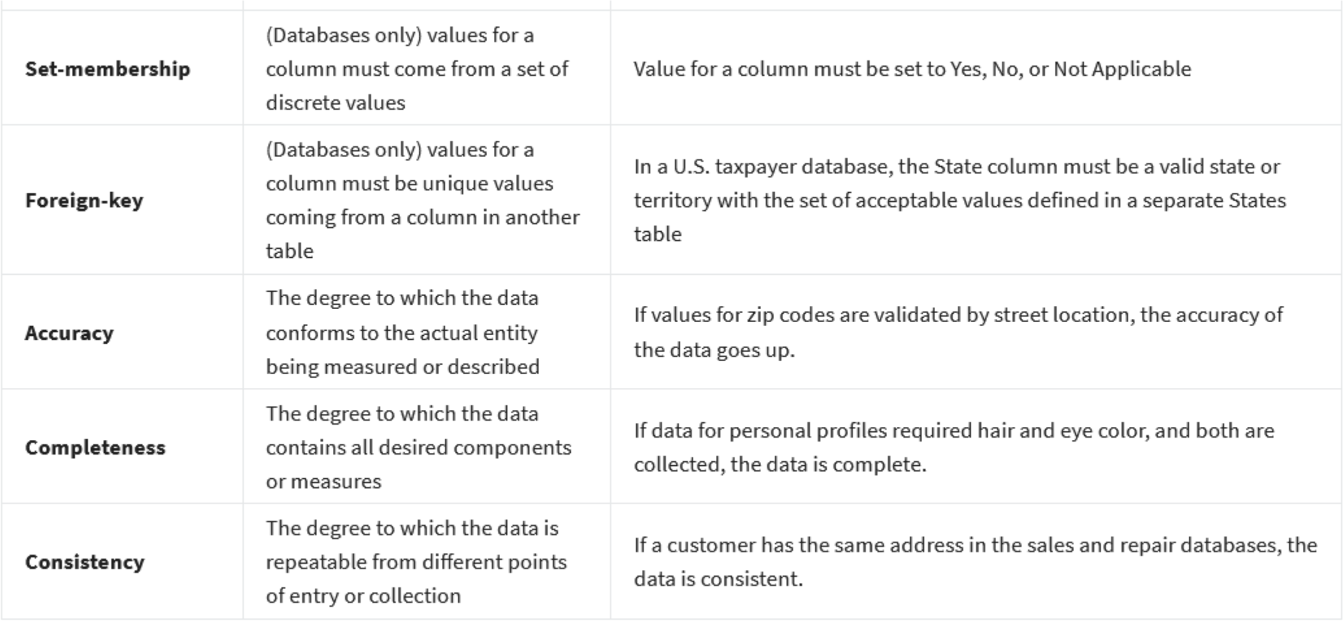
***Data replication compromising data integrity**: Continuing with the example, imagine you ask your international counterparts to verify dates and stick to one format. One analyst copies a large dataset to check the dates. But because of memory issues, only part of the dataset is actually copied. The analyst would be verifying and standardizing incomplete data. That partial dataset would be certified as compliant but the full dataset would still contain dates that weren't verified. Two versions of a dataset can introduce inconsistent results. A final audit of results would be essential to reveal what happened and correct all dates.*

***Data transfer compromising data integrity**: Another analyst checks the dates in a spreadsheet and chooses to import the validated and standardized data back to the database. But suppose the date field from the spreadsheet was incorrectly classified as a text field during the data import (transfer) process. Now some of the dates in the database are stored as text strings. At this point, the data needs to be cleaned to restore its integrity.*

***Data manipulation compromising data integrity**: When checking dates, another analyst notices what appears to be a duplicate record in the database and removes it. But it turns out that the analyst removed a unique record for a company's subsidiary and not a duplicate record for the company. Your dataset is now missing data and the data must be restored for completeness.*



Data constraint	Definition	Examples
Data type	Values must be of a certain type: date, number, percentage, Boolean, etc.	If the data type is a date, a single number like 30 would fail the constraint and be invalid
Data range	Values must fall between predefined maximum and minimum values	If the data range is 10-20, a value of 30 would fail the constraint and be invalid
Mandatory	Values can't be left blank or empty	If age is mandatory, that value must be filled in
Unique	Values can't have a duplicate	Two people can't have the same mobile phone number within the same service area
Regular expression (regex) patterns	Values must match a prescribed pattern	A phone number must match ###-###-#### (no other characters allowed)
Cross-field validation	Certain conditions for multiple fields must be satisfied	Values are percentages and values from multiple fields must add up to 100%
Primary-key	(Databases only) value must be unique per column	A database table can't have two rows with the same primary key value. A primary key is an identifier in a database that references a column in which each value is unique. More information about primary and foreign keys is provided later in the program.
Set-membership	(Databases only) values for a column must come from a set of discrete values	Value for a column must be set to Yes, No, or Not Applicable



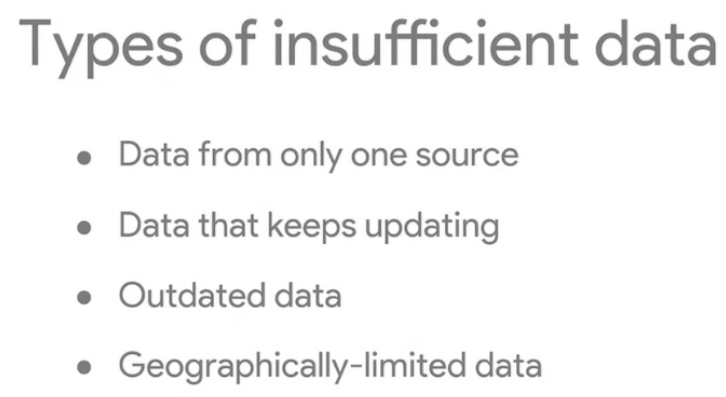
Set-membership	(Databases only) values for a column must come from a set of discrete values	Value for a column must be set to Yes, No, or Not Applicable
Foreign-key	(Databases only) values for a column must be unique values coming from a column in another table	In a U.S. taxpayer database, the State column must be a valid state or territory with the set of acceptable values defined in a separate States table
Accuracy	The degree to which the data conforms to the actual entity being measured or described	If values for zip codes are validated by street location, the accuracy of the data goes up.
Completeness	The degree to which the data contains all desired components or measures	If data for personal profiles required hair and eye color, and both are collected, the data is complete.
Consistency	The degree to which the data is repeatable from different points of entry or collection	If a customer has the same address in the sales and repair databases, the data is consistent.

Well-aligned objectives and data

Clean data + alignment to business objective = accurate conclusions

Alignment to business objective + additional data cleaning = accurate conclusions

Overcoming the challenges of insufficient data



Types of insufficient data

- Data from only one source
- Data that keeps updating
- Outdated data
- Geographically-limited data

Ways you can address them

- *You can identify trends with the available data*
- *wait for more data if time allows*
- *you can talk with stakeholders and adjust your objective*
- *you can look for a new data set*

What to do when you find an issue with your data

Data issue

No data:

- *Gather the data on a small scale to perform a preliminary analysis and then request additional time to complete the analysis after you have collected more data.*

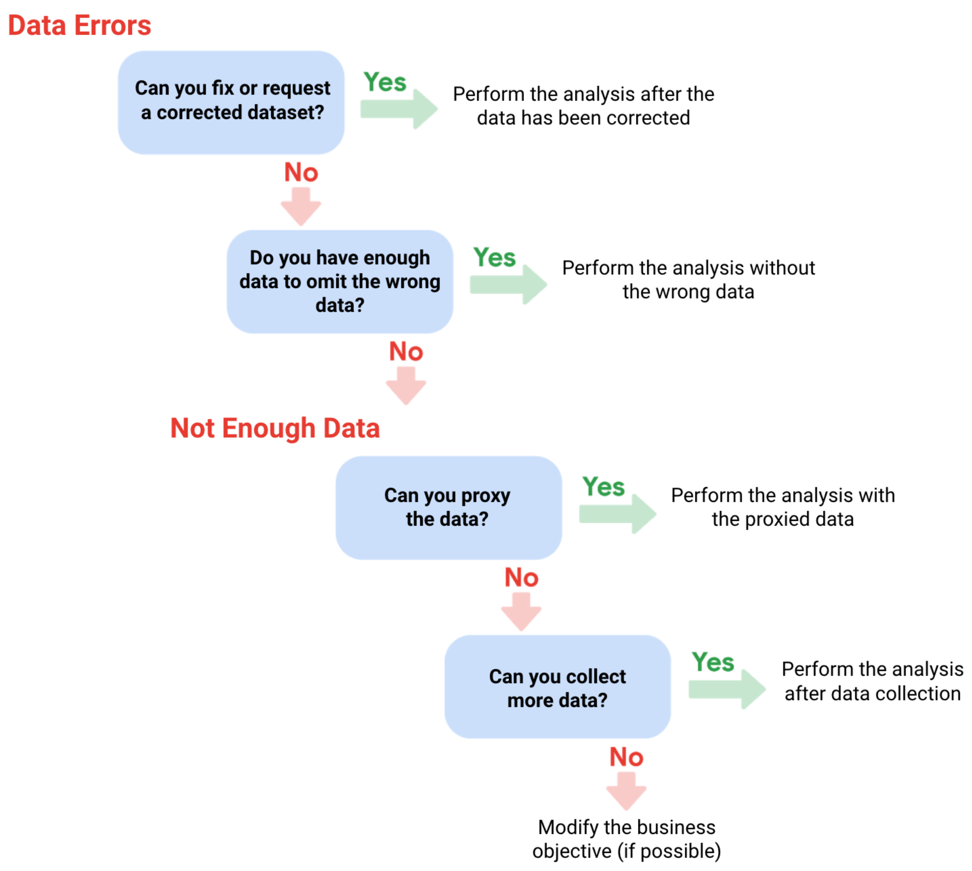
- *If there isn't time to collect data, perform the analysis using proxy data from other datasets. This is the most common workaround.*

Too little data:

- *Do the analysis using proxy data along with actual data.*
- *Adjust your analysis to align with the data you already have.*

wrong data, including data with errors*

- *If you have the wrong data because requirements were misunderstood, communicate the requirements again.*
- *Identify errors in the data and, if possible, correct them at the source by looking for a pattern in the errors.*
- *If you can't correct data errors yourself, you can ignore the wrong data and go ahead with the analysis if your sample size is still large enough and ignoring the data won't cause systematic bias.*


Data Errors
Can you fix or request a corrected dataset?
Yes
Perform the analysis after the data has been corrected
No
Do you have enough data to omit the wrong data?
Yes
Perform the analysis without the wrong data
No
Not Enough Data
Can you proxy the data?
Yes
Perform the analysis with the proxied data
No
Can you collect more data?
Yes
Perform the analysis after data collection
No
Modify the business objective (if possible)

The importance of sample size

Population is all possible data values in a certain dataset.

sample size!

When you use sample size or a sample, you use a part of a population that's representative of the population.

Sampling bias is when a sample isn't representative of the population as a whole. This means some members of the population are being overrepresented or underrepresented.

Random sampling is a way of selecting a sample from a population so that every possible type of the sample has an equal chance of being chosen.

Things to remember when determining the size of your sample

- Don't use a sample size less than 30.
- The confidence level most commonly used is 95%, but 90% can work in some cases.

Increase the sample size to meet specific needs of your project:

- For a **higher** confidence level, use a larger sample size
- To **decrease** the margin of error, use a larger sample size
- For **greater** statistical significance, use a larger sample size

Why a minimum sample of 30?

Central Limit Theorem (CLT)

Testing your data

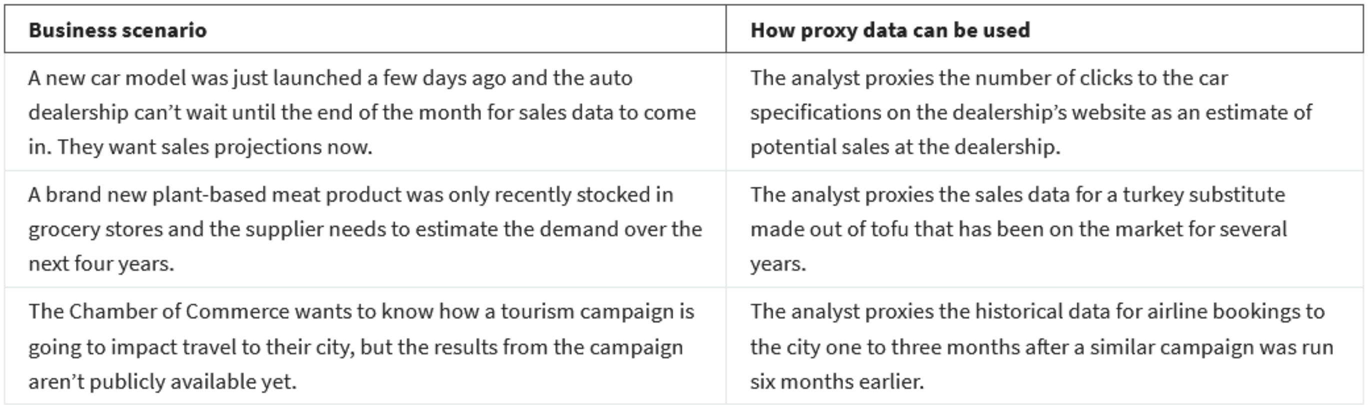
Using statistical power

Hypothesis testing

If a test is statistically significant,

It means the results of the test are real and not an error caused by random chance.

What to do when there is no data



Business scenario	How proxy data can be used
A new car model was just launched a few days ago and the auto dealership can't wait until the end of the month for sales data to come in. They want sales projections now.	The analyst proxies the number of clicks to the car specifications on the dealership's website as an estimate of potential sales at the dealership.
A brand new plant-based meat product was only recently stocked in grocery stores and the supplier needs to estimate the demand over the next four years.	The analyst proxies the sales data for a turkey substitute made out of tofu that has been on the market for several years.
The Chamber of Commerce wants to know how a tourism campaign is going to impact travel to their city, but the results from the campaign aren't publicly available yet.	The analyst proxies the historical data for airline bookings to the city one to three months after a similar campaign was run six months earlier.

open data is the information that has been published on government-sanctioned portals. In the best case, this data is structured, machine-readable, open-licensed, and well maintained.

Public data is the data that exists everywhere else. This is information that's freely available (but not really accessible) on the web. It is frequently unstructured and unruly, and its usage requirements are often vague.

Different types of data set on kaggle

- https://www.kaggle.com/datasnaek/youtube-new
- https://www.kaggle.com/sakshigoyal7/credit-card-customers
- https://www.kaggle.com/rtatman/188-million-us-wildfires
- https://www.kaggle.com/bigquery/google-analytics-sample

Sample size calculator

- https://www.surveymonkey.com/mp/sample-size-calculator/
- http://www.raosoft.com/samplesize.html

Consider the margin of error

Margin of error is the maximum amount that the sample results are expected to differ from those of the actual population.

Eg: Imagine you are playing baseball and that you are up at bat. The crowd is roaring, and you are getting ready to try to hit the ball. The pitcher delivers a fastball traveling about 90-95mph, which takes about 400 milliseconds (ms) to reach the catcher's glove. You swing and miss the first pitch because your timing was a little off. You wonder if you should have swung slightly earlier or slightly later to hit a home run. That time difference can be considered the margin of error, and it tells us how close or far your timing was from the average home run swing.

Week-2

Data cleaning is a must

Clean it up!

Dirty data is data that's incomplete, incorrect, or irrelevant to the problem you're trying to solve.
Clean data is data that's complete, correct, and relevant to the problem you're trying to solve.

What is dirty data?

- Types of dirty data you may encounter
- What may have caused the data to become dirty
- How dirty data is harmful to businesses

Types of dirty data

Duplicate data

Outdated data

Incomplete data

Incorrect/inaccurate data

Inconsistent data

Inconsistent data

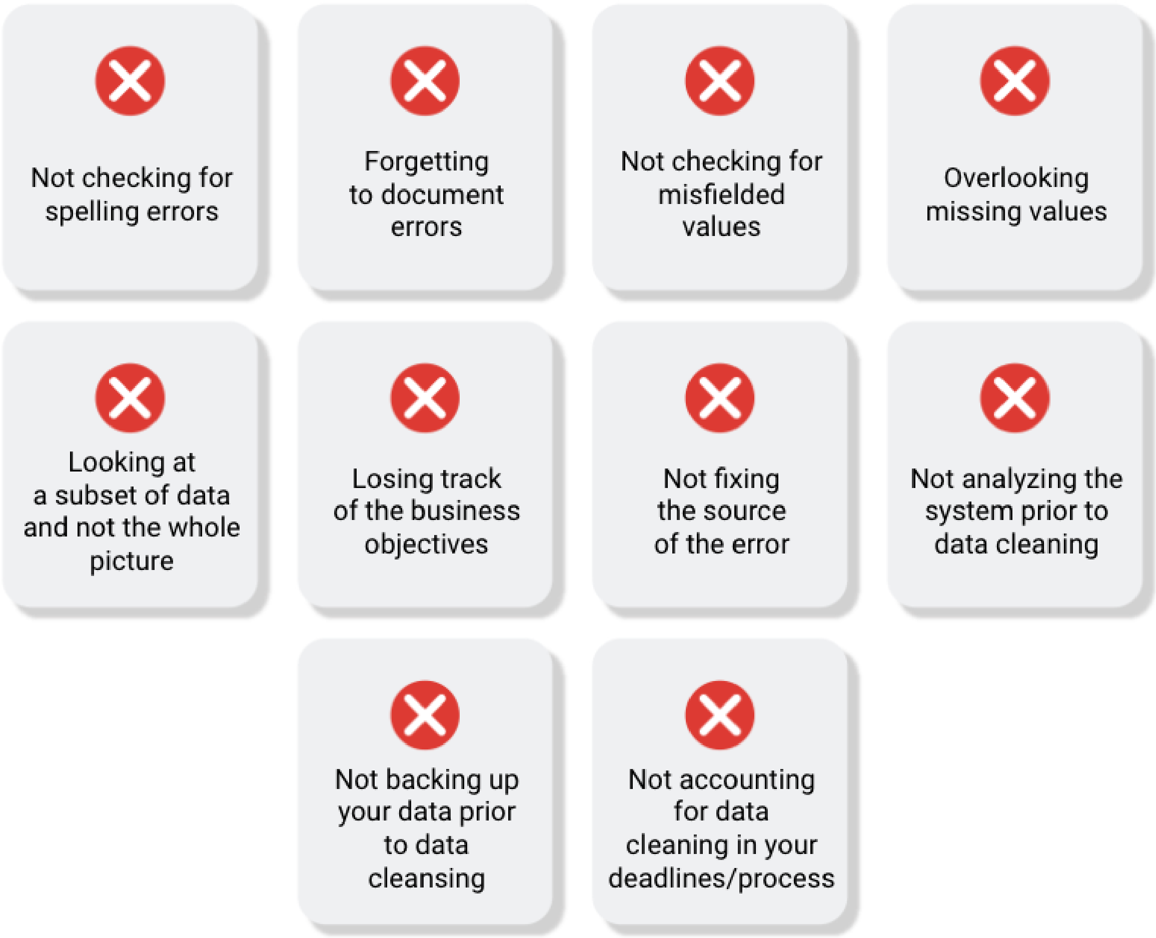
Any data that uses different formats to represent the same thing

Field *is a single piece of information from a row or column of a spreadsheet.*

Data validation *is a tool for checking the accuracy and quality of data before adding or importing it.*

Begin cleaning data

Common data-cleaning pitfalls



- Not checking for spelling errors
- Forgetting to document errors
- Not checking for misfielded values
- Overlooking missing values
- Looking at a subset of data and not the whole picture
- Losing track of the business objectives
- Not fixing the source of the error
- Not analyzing the system prior to data cleaning
- Not backing up your data prior to data cleansing
- Not accounting for data cleaning in your deadlines/process

Top ten ways to clean your data

- https://support.microsoft.com/en-us/office/top-ten-ways-to-clean-your-data-2844b620-677c-47a7-ac3e-c2e157d1db19
- https://support.google.com/a/users/answer/9604139?hl=en#zippy=

Hands-On Activity: Cleaning data with spreadsheets

- Filter
- Transpose (while pasting)
- Data cleanup (option: cleanup suggestion)
- Change text format (Add on: caps to lower etc…)

Cleaning data in spreadsheets

Data-cleaning features in spreadsheets

Conditional formatting

- Conditional formatting (to find empty cell)
- Remove duplicates
- Date formatting (format->number->date)
- specified text separating also called the delimiter.
- Data validation

Optimize the data-cleaning process

A function is a set of instructions that performs a specific calculation using the data in a spreadsheet.

Some basic types of functions in Spreadsheet

- COUNTIF
- LEN
- LEFT
- RIGHT
- CONCATENATE
- TRIM

Workflow automation

- https://towardsdatascience.com/automating-scientific-data-analysis-part-1-c9979cd0817e
- https://news.mit.edu/2016/automating-big-data-analysis-1021
- https://technologyadvice.com/blog/information-technology/top-10-workflow-automation-software/

Different data perspectives

- Pivot table
- VLOOKUP - vertical lookup
- Find
- Graph plotting

Even more data-cleaning techniques

Data mapping is the process of matching fields from one database to another.

Compatibility describes how well two or more data sets are able to work together.

- CONCATENATE

Hands-On Activity: Clean data with spreadsheet functions

- SPLIT
- COUNTIF
- Sort

Learning Log: Develop your approach to cleaning data

Step 1: Create your checklist

Some things you might include in your checklist:

- Size of the data set
- Number of categories or labels
- Missing data
- Unformatted data
- The different data types

Step 2: List your preferred cleaning methods

After you have compiled your personal checklist, you can create a list of activities you like to perform when cleaning data. This list is a collection of procedures that you will implement when you encounter specific issues *present in the data related to your checklist or every time you clean a new dataset.*

For example, suppose that you have a dataset with missing data, how would you handle it? Moreover, if the data set is very large, what would you do to check for missing data? Outlining some of your preferred methods for cleaning data can help save you time and energy.

Step 3: Choose a data cleaning motto

Now that you have a personal checklist and your preferred data cleaning methods, you can create a data cleaning motto to help guide and explain your process. The motto is a short one or two sentence summary of your philosophy towards cleaning data. For example, here are a few data cleaning mottos from other data analysts:

1. *"Not all data is the same, so don't treat it all the same."*
2. *"Be prepared for things to not go as planned. Have a backup plan."*
3. *"Avoid applying complicated solutions to simple problems."*

My list

- *Find Empty cell*
- *Remove duplicates*
- *Date format*
- *Split wanted informations*
- *Check conditions*

Week - 3

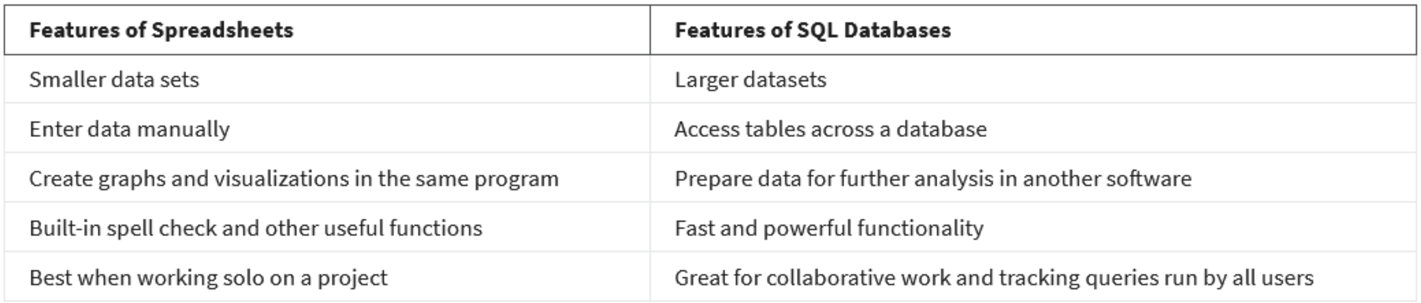
Using SQL to clean data

Understanding SQL capabilities

Relational databases

This is a database that contains a series of tables that can be connected to form relationships.

Using SQL as a junior data analyst



Features of Spreadsheets	Features of SQL Databases
Smaller data sets	Larger datasets
Enter data manually	Access tables across a database
Create graphs and visualizations in the same program	Prepare data for further analysis in another software
Built-in spell check and other useful functions	Fast and powerful functionality
Best when working solo on a project	Great for collaborative work and tracking queries run by all users

SQL dialects and their uses

Links to learn SQL documentation

- https://learnsql.com/blog/what-sql-dialect-to-learn/
- https://www.softwaretestinghelp.com/sql-vs-mysql-vs-sql-server/
- https://www.datacamp.com/community/blog/sql-differences
- https://sqlite.org/windowfunctions.html
- https://www.sqltutorial.org/what-is-sql/

Hands-On Activity: Processing time with SQL

```
SELECT
  language,
  title,
  SUM(views) AS views
FROM
  `bigquery-samples.wikipedia_benchmark.Wiki10B`
WHERE
  title LIKE '%Google%'
GROUP BY
  language,
  title
ORDER BY
  views DESC;
```

Learn basic SQL queries

Widely used SQL queries

- INSERT INTO
- VALUES
- UPDATE
- SET
- SELECT   COUNT SUM * DISTINCT
- FROM
- WHERE
- ORDERED BY
- GROUP BY
- LIMIT

SELECT

- COUNT
- SUM
- *
- DISTINCT
- LENGTH()

WHERE

- SUBSTR()
- TRIM()
- LENGTH()

Hands-On Activity: Clean data using SQL

- *MIN*
- *MAX*
- *UPDATE*
- *SET*
- *DISTINCT*

Step 1 :

```
SELECT
    DISTINCT(fuel_type)
FROM
    `dulcet-velocity-294320.From_course.automobile_data`
Multi line command
*/
--STEP 2
/*SELECT
  MIN(length) as min_length,
  MAX(length) as max_length
FROM
  `dulcet-velocity-294320.From_course.automobile_data`*/



--STEP 3
/*SELECT
  *
FROM
  `dulcet-velocity-294320.From_course.automobile_data`
```

```
WHERE
    num_of_doors is NULL*/


--STEP 4
/*UPDATE
  `dulcet-velocity-294320.From_course.automobile_data`
SET
  num_of_doors = "four"
WHERE
  make = "dodge"
  AND fuel_type = "gas"
  AND body_style = "sedan";*/  --KASU KATUNA THA WORK AAGUM


--MY CODE
/*SELECT
    *
FROM
`dulcet-velocity-294320.From_course.automobile_data`
WHERE
    make = "dodge"
    OR fuel_type = "gas"
    AND body_style = "sedan"*/

--STEP 5
/*SELECT
```

```
    DISTINCT(num_of_cylinders)
FROM
    `dulcet-velocity-294320.From_course.automobile_data`*/


--STEP 6
/*UPDATE
  cars.car_info
SET
  num_of_cylinders = "two"
WHERE
  num_of_cylinders = "tow";*/


--STEP 7
/*SELECT
  MIN(compression_ratio) AS min_compression_ratio,
  MAX(compression_ratio) AS max_compression_ratio
FROM
  `dulcet-velocity-294320.From_course.automobile_data`
WHERE
        compression_ratio <> 70;*/  --omit 70



--STEP 8
/*SELECT
```

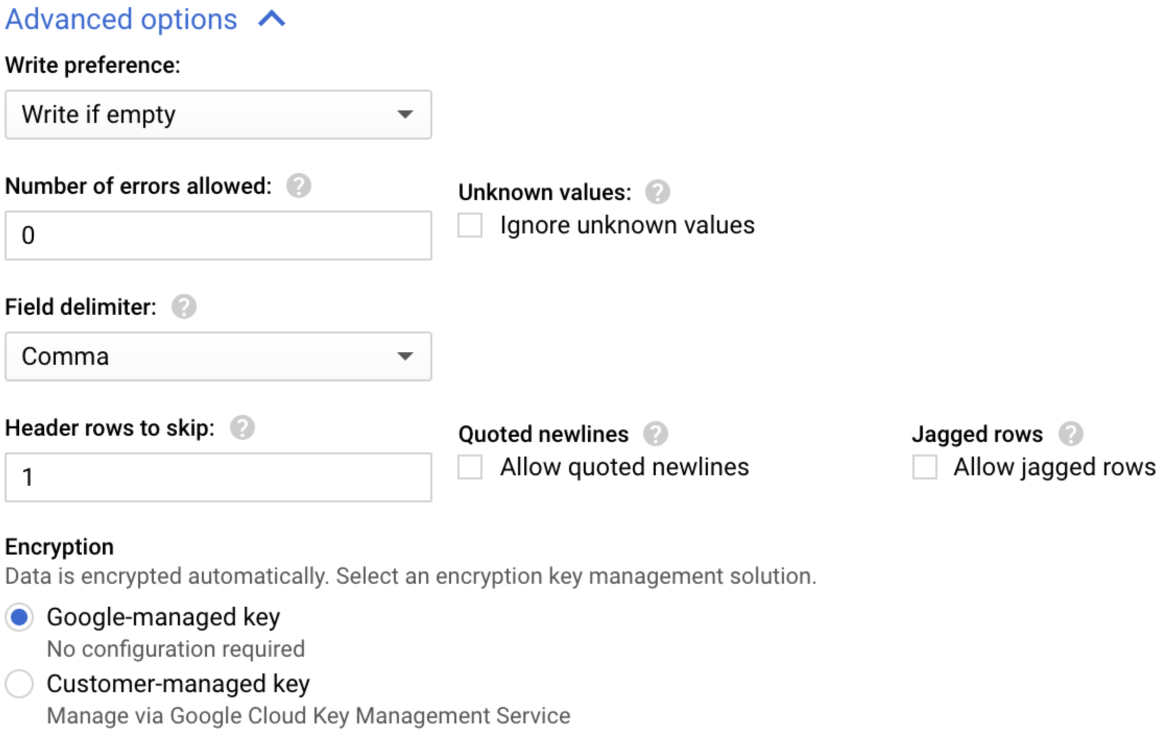
```
    COUNT(*) AS num_of_rows_to_delete
FROM
    `dulcet-velocity-294320.From_course.automobile_data`
WHERE
    compression_ratio = 70;*/


--STEP 9
/*DELETE
    `dulcet-velocity-294320.From_course.automobile_data`
WHERE
    compression_ratio = 70;*/


--STEP 9
/*SELECT
  DISTINCT drive_wheels,
  LENGTH(drive_wheels) AS string_length
FROM
  `dulcet-velocity-294320.From_course.automobile_data`*/




--STEP 10
/*UPDATE
    cars.car_info
```

```
SET
    drive_wheels = TRIM(drive_wheels)
WHERE
         TRUE;*/



--STEP 10
/*SELECT
    TRIM(drive_wheels),
    LENGTH(drive_wheels) AS string_length,
FROM
    `dulcet-velocity-294320.From_course.automobile_data`*/

--TEST
/*SELECT
    MAX(price) as MAX_PRICE
FROM
    `dulcet-velocity-294320.From_course.automobile_data`*/
```

Transforming data

Upload the store transactions dataset to BigQuery

```
[
{
"description": "date",
"mode": "NULLABLE",
"name": "date",
 "type": "DATETIME"
},
 {
"description": "transaction id",
"mode": "NULLABLE",
 "name": "transaction_id",
 "type": "INTEGER"
},
{
 "description": "customer id",
 "mode": "NULLABLE",
"name": "customer_id",
"type": "INTEGER"
 },
 {
"description": "product name",
 "mode": "NULLABLE",
"name": "product",
 "type": "STRING"
 },
 {
 "description": "product_code",
"mode": "NULLABLE",
 "name": "product_code",
 "type": "STRING"
},
{
"description": "product color",
 "mode": "NULLABLE",
"name": "product_color",
 "type": "STRING"
 },
 {
 "description": "product price",
 "mode": "NULLABLE",
"name": "product_price",
 "type": "FLOAT"
},
{
"description": "quantity purchased",
"mode": "NULLABLE",
"name": "purchase_size",
 "type": "INTEGER"
 },
{
"description": "purchase price",
 "mode": "NULLABLE",
 "name": "purchase_price",
"type": "STRING"
 },
{
 "description": "revenue",
 "mode": "NULLABLE",
"name": "revenue",
 "type": "FLOAT"
 }
]
```



Advanced options

Write preference: Write if empty

Number of errors allowed: 0

Unknown values: □ Ignore unknown values

Field delimiter: Comma

Header rows to skip: 1

Quoted newlines: □ Allow quoted newlines

Jagged rows: □ Allow jagged rows

Encryption

Data is encrypted automatically. Select an encryption key management solution.

● Google-managed key

No configuration required

○ Customer-managed key

Manage via Google Cloud Key Management Service

Three types of file uploading

- Direct csv file upload which has header
- Txt file upload with including headers and type, same for csv which does not have header
- Changing data type while uploading with headers

Type conversion

PART-1

```
SELECT
    *
FROM
    `dulcet-velocity-294320.From_course.customer_purchase`
ORDER BY
    CAST(purchase_price AS FLOAT64 ) DESC
```

PART-2

```
--SORTING WITH DATE
/*
SELECT
    date,
    purchase_price
FROM
    `dulcet-velocity-294320.From_course.customer_purchase`
WHERE
    date BETWEEN '2020-12-1' and '2020-12-31' */
```

```
--CAST Change data types
/*
SELECT
    CAST(date as date) as DATE,
    purchase_price
FROM
    `dulcet-velocity-294320.From_course.customer_purchase`
ORDER BY
    CAST(date as date) */



--CONCAT join strings to form substring
/*
SELECT
    CONCAT(product_code,product_color) as unic_clr_id
FROM
    `dulcet-velocity-294320.From_course.customer_purchase`
WHERE
    product = 'couch'*/


--COALESCE() return non null values
SELECT
    COALESCE(product,product_code) as product_info
FROM
    `dulcet-velocity-294320.From_course.customer_purchase`
```

Part 1 & 2

CAST()	change data type
CONCAT()	join 2 string
COALESCE()	from this or that

Week - 4

Manually cleaning data

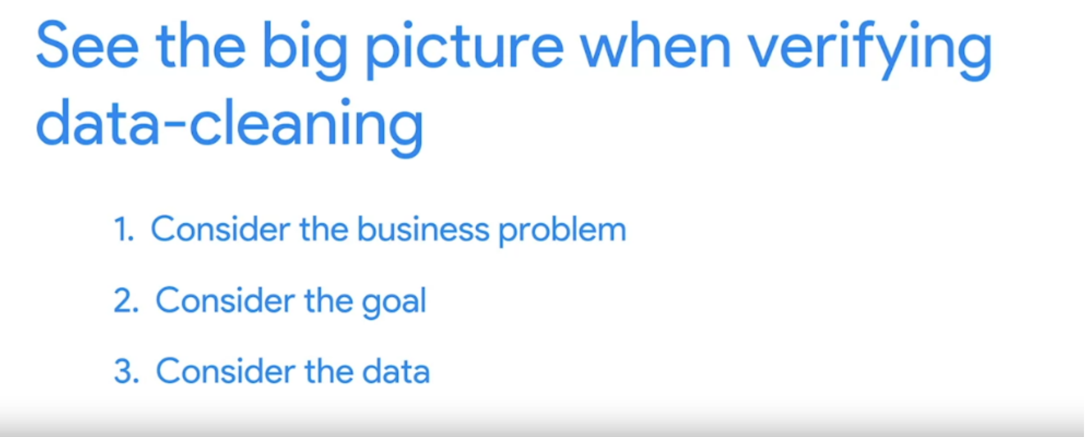
Verifying and reporting results

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

A changelog is a file containing a chronologically ordered list of modifications made to a project.

Cleaning and your data expectations

- Using spreadsheet
- Use SQL
- Big picture verification (including graphs)



See the big picture when verifying data-cleaning

1. Consider the business problem
2. Consider the goal
3. Consider the data

The final step in data cleaning

- *Spell check*
- *Spreadsheet => find and replace*
- *SQL => CASE*

```
SELECT
    customer_id,
    CASE
        WHEN product = 'fan' THEN 'FAN'
        WHEN product = 'lamps' THEN 'LAMP'
        ELSE product
        END AS Dhamu
FROM `dulcet-velocity-294320.From_course.customer_purchase`
```

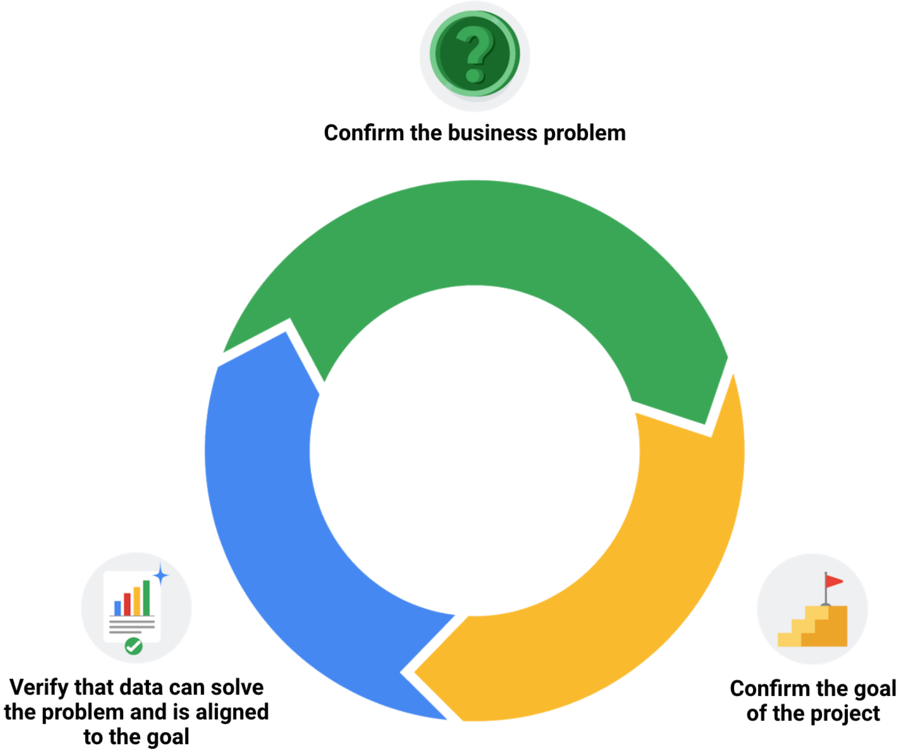
Data-cleaning verification: A checklist

- ***Sources of errors**: Did you use the right tools and functions to find the source of the errors in your dataset?*
- ***Null data**: Did you search for NULLs using conditional formatting and filters?*
- ***Misspelled words**: Did you locate all misspellings?*
- ***Mistyped numbers**: Did you double-check that your numeric data has been entered correctly?*
- ***Extra spaces and characters**: Did you remove any extra spaces or characters using the **TRIM** function?*
- ***Duplicates**: Did you remove duplicates in spreadsheets using the **Remove Duplicates** function or **DISTINCT** in SQL?*
- ***Mismatched data types**: Did you check that numeric, date, and string data are typecast correctly?*

- ***Messy (inconsistent) strings***: *Did you make sure that all of your strings are consistent and meaningful?*
- ***Messy (inconsistent) date formats***: *Did you format the dates consistently throughout your dataset?*
- ***Misleading variable labels (columns)***: *Did you name your columns meaningfully?*
- ***Truncated data***: *Did you check for truncated or missing data that needs correction?*
- ***Business Logic***: *Did you check that the data makes sense given your knowledge of the business?*

The goal of your project

- *Confirm the business problem*
- *Confirm the goal of the project*
- *Verify that data can solve the problem and is aligned to the goal*


Confirm the business problem
Verify that data can solve the problem and is aligned to the goal
Confirm the goal of the project

Documenting results and the cleaning process

Capturing cleaning changes

Documentation

Which is the process of tracking changes, additions, deletions and errors involved in your data cleaning effort.

Data errors are the crime, data cleaning is gathering evidence, and documentation is detailing exactly what happened for peer review or court.

Here is how a version control system affects a change to a query:

1. *A company has official versions of important queries in their **version control system**.*
2. *An analyst makes sure the most up-to-date version of the query is the one they will change. This is called **syncing***
3. *The analyst makes a change to the query.*
4. *The analyst might ask someone to review this change. This is called a **code review** and can be informally or formally done. An informal review could be as simple as asking a senior analyst to take a look at the change.*
5. *After a reviewer approves the change, the analyst submits the updated version of the query to a repository in the company's version control system. This is called a **code commit**. A best practice is to document exactly what the change was and why it was made in a comments area. Going back to our example of a query that pulls daily revenue, a comment might be: Updated revenue to include revenue coming from the new product, Calypso.*

6. *After the change is **submitted**, everyone else in the company will be able to access and use this new query when they **sync** to the most up-to-date queries stored in the version control system.*
7. *If the query has a problem or business needs change, the analyst can **undo** the change to the query using the version control system. The analyst can look at a chronological list of all changes made to the query and who made each change. Then, after finding their own change, the analyst can **revert** to the previous version.*
8. *The query is back to what it was before the analyst made the change. And everyone at the company sees this reverted, original query, too.*
9. *Changelogs are for humans, not machines, so write legibly.*

Embrace changelogs

In

- *Spreadsheet*
- *Excel*
- *Big query*

Typically, a changelog records this type of information:

- *Data, file, formula, query, or any other component that changed*
- *Description of what changed*
- *Date of the change*
- *Person who made the change*
- *Person who approved the change*
- *Version number*
- *Reason for the change*

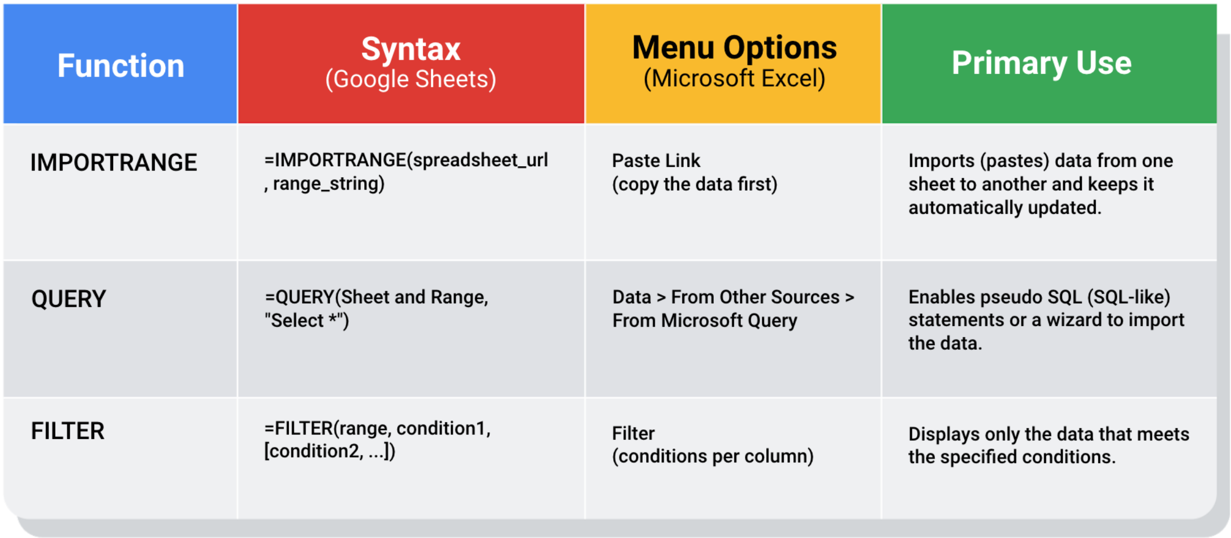
Changelog documentation

```
# Changelog

This file contains the notable changes to the project


Version 1.0.0 (02-23-2019)

## New

    - Added column classifiers (Date, Time, PerUnitCost, TotalCost, etc. )

    - Added Column “AveCost” to track average item cost


## Changes

    - Changed date format to MM-DD-YYYY

    - Removal of whitespace (cosmetic)


## Fixes

    - Fixed misalignment in Column "TotalCost" where some rows did not
match with correct dates

    - Fixed SUM to run over entire column instead of partial
```

Some of the most common errors involve

- *human mistakes like mistyping or misspelling,*
- *flawed processes like poor design of a survey form, and*
- *system issues where older systems integrate data incorrectly.*

Advanced functions for speedy data cleaning



Function	Syntax (Google Sheets)	Menu Options (Microsoft Excel)	Primary Use
IMPORTRANGE	=IMPORTRANGE(spreadsheet_url , range_string)	Paste Link (copy the data first)	Imports (pastes) data from one sheet to another and keeps it automatically updated.
QUERY	=QUERY(Sheet and Range, "Select *")	Data > From Other Sources > From Microsoft Query	Enables pseudo SQL (SQL-like) statements or a wizard to import the data.
FILTER	=FILTER(range, condition1, [condition2, ...])	Filter (conditions per column)	Displays only the data that meets the specified conditions.

- *QUERY*
- *IMPORTRANGE*
- *FILTER*

QUERY

- *https://support.google.com/docs/answer/3093343?hl=en*

Filter

- *https://support.google.com/docs/answer/3093197?hl=en*
- *https://support.google.com/docs/answer/3093197?hl=en*

IMPORTRANGE

- *https://support.google.com/docs/answer/3093340?hl=en#*

Week - 5

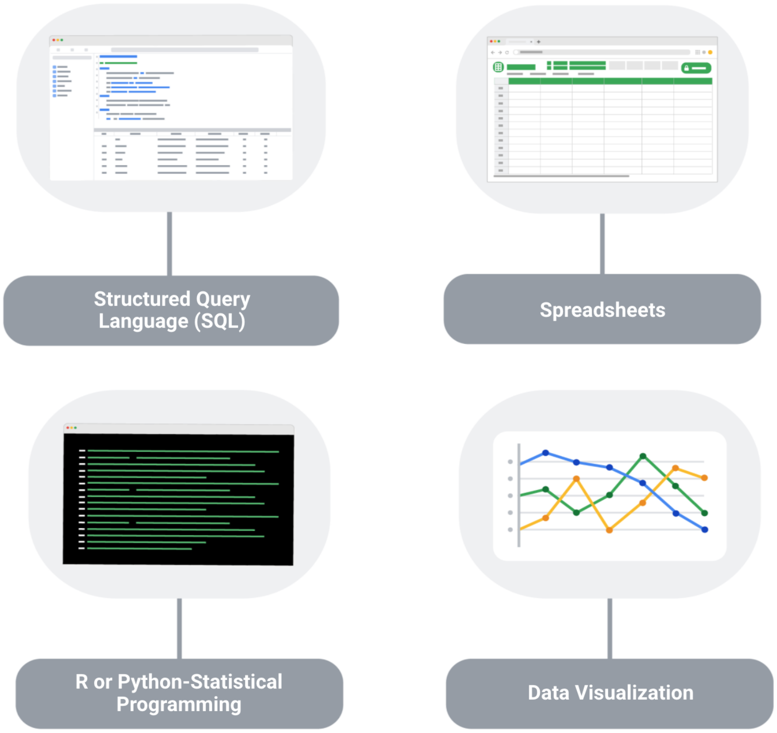
Understand the elements of a data analyst resume

CareerCon resources on YouTube

Youtube links

- https://www.youtube.com/playlist?list=PLqFaTIg4myu-npFrYu6cO7h7AI6bkcOlL
- https://www.youtube.com/watch?v=cBbYhhH399c&list=PLqFaTIg4myu-npFrYu6cO7h7AI6bkcOlL&index=9

Adding professional skills to your resume


Structured Query Language (SQL)
Spreadsheets
R or Python-Statistical Programming
Data Visualization

Highlighting experiences on resumes

Adding soft skills to your resume

1. Presentation Skills
2. Collaboration
3. Communication
4. Research
5. Problem-solving skills
6. Adaptability
7. Attention to detail

Quick Review

Week -1

- Data integrity
- Manage insufficient data
- Statistics

Week - 2

- Spreadsheet

Week - 3

- SQL

Week - 4

- Verification and Cleaning
- Changelog and documentation
- Checklist

Week - 5

- Hiring process
- Resume building

DHAMODHARAN
14/10/2021