Google data analytics

Data generation --> Anything one does online with smart devices. There are more than 1.2trn serach per year and more than 2bn users on youtube.
Data analyst = Someone who collects, transforms, and organizes data in order to help make informed decisions.
Data--> A collection of facts.
Can't make brics without clay!.... Sherlock Holmes via Sir Arthur Conan Doyle.
Data analytics = The science of data. [Broader concept]
Data analysis = Creating insights from data. [narrower concept]
Data analysis process = ask, prepare, process, analyze, share, and act. The SAS model emphasizes the cyclical nature of their model by visualizing it as an infinity symbol. + Gut instinct [balance these like detective does].
Career identity statement includes =
Your strengths = what makes you feel strengthened. [Research + motivate others]
your motivations = What passionates you [New insights and ideas + share them + cracking jokes]
Your values = plans implementation + punctual + family
—----------

History = Statistics --> Egyptian pyramids + documented their calculations and theories on papyri (paper-like materials), which are now viewed as the earliest examples of spreadsheets and checklists.
Dell EMC Corporation's data analytics process is cyclical with six steps: (David Dietrich)
Discovery
Pre-processing data
Model planning
Model building
Communicate results
Operationalize

Data ecosystems --> The various elements that interact with one another in order to produce, manage, store, organisze, analyze, and share data.
Data can also be found on cloud.
Cloud = A place to keep data online, rather thsn on a computer hard drive.

Data science = Creating new ways of modeling and understanding the unknown by using raw data. [Find new questions vs analysts find answers / new insights]

D analytics
D analysis

Data driven decision making --> Rise of e-commerce [Jeff Bezos = data showed 90% hike in books ordering]
Analytical skills = solving poblems with facts. To review, these skills are:
Curiosity
Understanding of context
Technical mindset
Data design
Data strategy

Analytical thinking =
Correlation is not equal to causation.
Gap analysis
5 Why technique**

Data life cycle [Caterpillar / electricity provider] =
Plan --> Capture --> Manage --> analyze --> Archive --> Destroy
Databases --> A collection of data stored in a computer system. [data integrity and ethics]
Data integrity → Data integrity is the process of ensuring that an organization's data is accurate, complete, and consistent. It's important to maintain data integrity to protect sensitive information, ensure reliable data analytics, and comply with regulatory frameworks.
—---
Algorithm = A process or set of rules to be followed for a specific task.
Data --> Information (context + comparison) --> Knowledge

Metrics --> ROI;
Metric goal --> A measurable goal set by a company and evaluated using metrics.

Data visualization --> Florence Nightingale = Philosophy of modern nursing; Crimea war of 1870s --> Used charts = no. of deaths in several months --> Larger blue dots = Preventable deaths --> Changed the way of hospital care and nursing.

—----------

Functions vs Formulas → a **formula** is created by the **user**, while a **function** is a **predefined** calculation

—----------

A **statement of work** is a document that **clearly identifies the products and services** a vendor or contractor will provide to an organization. It includes objectives, guidelines, deliverables, schedule, and costs. SUPERSET.
A **scope of work** is project-based and **sets the expectations and boundaries of a project.** A scope of work may be included in a statement of work to help define project outcomes.
As a junior data analyst, it's more typical to be asked to create a scope of work than a statement of work.

—----------

https://dataedo.com/blog/basic-data-modeling-techniques
3 Basic **Data Modeling** Techniques - ERD, UML and Data Dictionary
1. Entity Relationship Diagrams--> default technique for modeling and the design of relational (traditional) databases. In this notation architect identifies:

Entities representing objects (or tables in relational database),
Attributes of entities including data type,
Relationships between entities/objects (or foreign keys in a database).
ERDs work well if you want to design a relational (classic) database, Excel databases or CSV files. Basically, any kind of tabular data. They work well for visualization of database schemas and communication of top-level view of data.
2. UML Class Diagrams--> standard for software engineering. It comprises of several different diagrams representing different aspect of the system, and one of them being a Class Diagram that can be used for data modeling. used to design classes in object-oriented programming languages (such as Java or C#).
3. Data Dictionary--> Techniques mentioned above were visual and were based on diagrams, and data dictionaries are a tabular definition/representation of data assets. Data dictionary is an inventory of data sets/tables with the list of their attributes/columns. Data dictionary is suitable as detailed specification of data assets and can be supplemented with ER diagrams, as both serve slightly different purpose.
—----------

Good data --> ROCCC--> Reliable / Original / Comprehensive / Current / Cited
—----------

Personally identifiable information, or PII, is information that can be used by itself or with other data to track down a person's identity. Data anonymization is the process of protecting people's private or sensitive data by eliminating that kind of information.
through de-identification, which is a process used to wipe data clean of all personally identifying information.

Ethical AI --> Democratise use use of AI + Evlolve AI ethically.

metadata is data about data. Photo / email ki description.
Descriptive --> ISBN no. of a book + author + title
Structural metadata --> How a piece of data is organised its relation with other data collection. E.g. chapters organiesed in a book + original version.
Administrative metadata --> Time photo was taken; png/jpeg; etc. Basically file properties.

CSV = Comma Separated Values

—----------
Statistical power can be calculated and reported for a completed experiment to comment on the confidence one might have in the conclusions drawn from the results of the study. It can also be used as a tool to estimate the number of observations or sample size required in order to detect an effect in an experiment." .8 or 80%
Margin of error --> Trust On results predicted on sample vs total population. → 60% +-10% [max variation]--> 50-70% yes
Confidence level --> 90-95% reliability on data → implies that 90-95% population is 50-70% yes
Sample size --> Population size ^ confidence level ^ margin of error

—----------
Data cleaning motto → “**Purge the noise, polish the truth”.**
—----------
Spreadsheet
VLOOKUP
COUNTIF
LEFT
RIGHT
MID
CONCATENATE
TRIM
Conditional FORMATTING
UPPER CASING add on
FILTER
PIVOT TABLE

VLOOKUP -->
a-b --> a | b into separate columns --> Data=> split text to columns --> manually "-"
OR use SPLIT

SQL

```
INSERT INTO "TABLE"
     (ID, CITY,STATE)
VALUES
     (2645, JAIPUR,RAJASATHAN)


UPDATE "TABLE"
SET address = '123 New address'
WHERE ID = '2645'

CREATE TABLE
```

—---

LENGTH / SUBSTR / TRIM

```
SELECT
  LENGTH (COUNTRY) AS LETTERS_IN_COUNTRY
FROM
  CUSTOMER_DATA.CUSTOMER_ADDRESS


SELECT
    DISTINCT customer_id
FROM
     customer_data.customer_address
WHERE
      SUBSTR(country,1,2) = 'US'

SELECT
  DISTINCT customer_id
FROM
   customer_data.customer_address
where
   TRIM (state) = 'OH'
```

—---

TYPECAST strings into FLOAT via CAST

```
SELECT
  CAST(purchase_price AS FLOAT64)
 FROM `direct-raceway-446106-u2.customer_data.customer_purchase`
```

```
ORDER BY
 CAST(purchase_price AS FLOAT64) DESC
```

Date **CAST**

```
SELECT
  CAST(date AS date) AS date_only,
  purchase_price
 FROM `direct-raceway-446106-u2.customer_data.customer_purchase`
 WHERE
 date BETWEEN '2020-12-01'AND '2020-12-31'
```

CONCAT

```
SELECT
    CONCAT (product_code, product_color) AS new_product_code
 FROM `direct-raceway-446106-u2.customer_data.customer_purchase`
 WHERE
    product = 'couch'
```

COALESCE --> Retreive non-null values

```
SELECT
    COALESCE(product, product_code) AS product_info
FROM `direct-raceway-446106-u2.customer_data.customer_purchase`
```

Your SQL experience (so far)

So far, you have been introduced to many different tools available in SQL. As a brief review, you learned how to complete tasks such as:

- Getting data from a table using **SELECT** statements.
- De-duplicating data using commands like **DISTINCT** and **COUNT + WHERE**.
- Manipulating string data with **TRIM()** and **SUBSTR**.
- Creating/dropping tables with **CREATE TABLE** and **DROP TABLE**.
- Changing data types with **CAST**.

CASE

```
SELECT

    customer_name,

    order_amount,

    CASE
```

```
        WHEN order_amount < 100 THEN "Small Order"
        WHEN order_amount BETWEEN 100 AND 500 THEN "Medium Order"
        ELSE "Large Order"
    END AS order_size
FROM orders;
```

```
SELECT
    cutomer_id,
    CASE
    WHEN first_name = 'Tnoy' THEN 'Tony'
    WHEN first_name = 'Tmoy' THEN 'Tony'
    ELSE first_name = 'Rachael' THEN 'Rachel'
    END AS cleaned_name
FROM customer_data.customer_name
```

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.

Resume ->

Selected as one of 275 participants nationwide for this 12-month professional development program for high-achieving talent based on leadership potential and academic success

Google certificate in resume intro

Entry-level data analytics professional; recently completed the Google Data Analytics Professional Certificate

---

Hard-working customer service representative with over five years of experience

1. *__Clear Communicator__
2. Summary

Transitioning from a career in the auto industry and seeking a full-time role in the field of data analytics

2. Work experience as *PAR --> Problem-Action-Result => **PROBLEM SOLVING** SKILLS

~~Was responsible for writing two blogs a month~~
Earned little-known website over 2,000 organic clicks through strategic blogging

Past work ex → Backed by data

Effectively implemented and communicated daily workflow procedures to fellow team members, resulting in a 15% increase in productivity

Soft skills → Detail oriented + Perseverance + Attention to Detail

How to better know yourself -->
How would you describe me to someone else?
What stands out about me?
How do I inspire you?

Career identity statement -->
I am a (role(s)) with (# years) of experience doing (accomplishment).
My greatest strength is (strength), and I have a talent for (strength). I am passionate about (motivation), and I value (value).