Data Analysis Fundamentals

Lecture 3: Understanding Data Operations

Vikas Kumar
Indian Institute of Technology Bombay

07th September 2023

Outline

- Recap
- Data Operations
- Select, Filter - Simple vs Complex, Sort, Group & Aggregate
- Merge, Pivot, Unpivot, Windowing

Recap

- Data and its evolution
- Data Science as a career option and typical job roles
- Data Science Pipeline and toolkit of a Data Analyst
- Excel and its history
- Data, Binning and Granularity
- Metrics and KPIs

Data Operations

- Processes applied to the data
- Processes can be mathematical/non-mathematical

=SUM(A1:A5)
=COUNT(A1:A5)
=AVERAGE(A1:A5)
=MIN(A1:A5)
=MAX(A1:A5)
=IF(A1>33,"P","F")

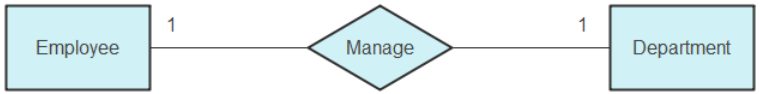
Dimensions and Facts

- A dimension is a measure of a physical variable (without numerical values).
- A unit is a way to assign a number or measurement to that dimension.
- For example, length is a dimension, but it is measured in units of feet (ft) or meters (m).
- A numeric attribute for which data operations can be done is called fact.
- A fact is the numerical part of the dimension.

Cardinality

- Cardinality means how the entities are arranged to each other.
- The relationship structure between entities in a relationship set.

one-to-one (1:1)

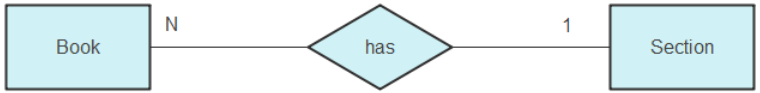


✓ One student can have only one student ID

one-to-many (1:N)

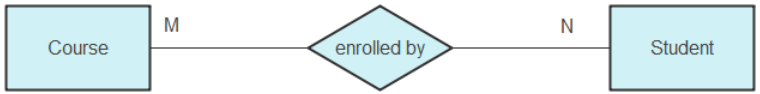
✓ Edyoda has multiple trainers

many-to-one (N:1)



✓ One trainer teaches multiple students in the class

many-to-many (M:N)



✓ One trainer takes several modules, and same module is taken by different trainers

Filter

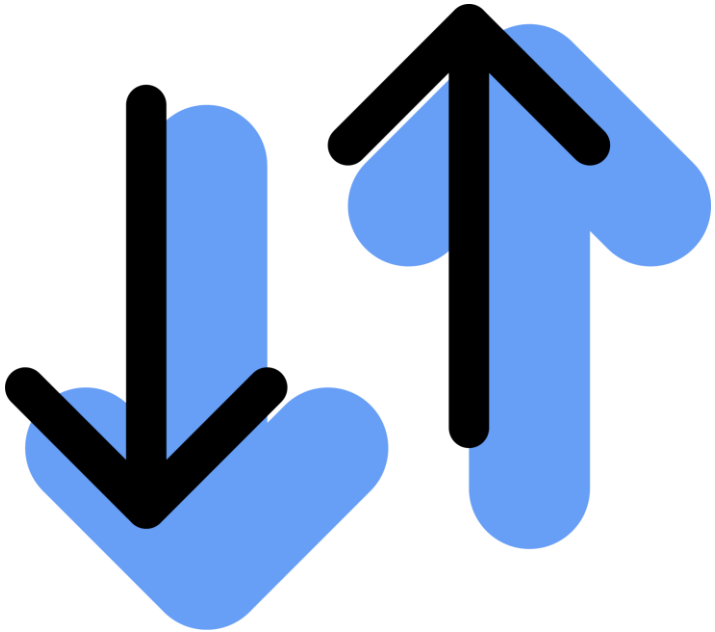
- Data filtering is the process of examining a dataset to exclude, rearrange, or apportion data according to certain criteria.

- Process of choosing a smaller part of your dataset and using that subset for viewing or analysis.

Sort

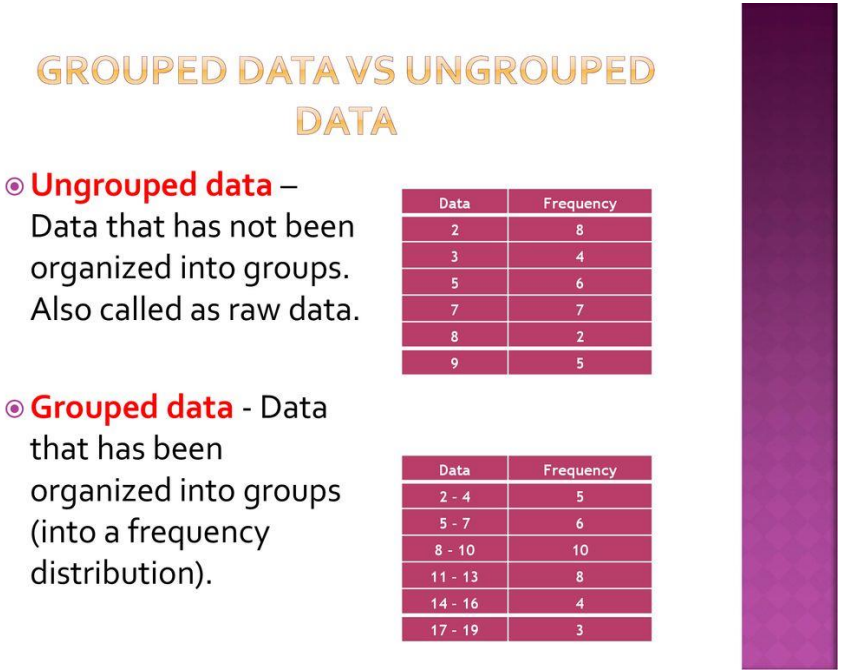
- Data sorting is any process that involves arranging the data into some meaningful order to make it easier to understand, analyze or visualize.

- Sorting refers to ordering data in an increasing or decreasing manner.



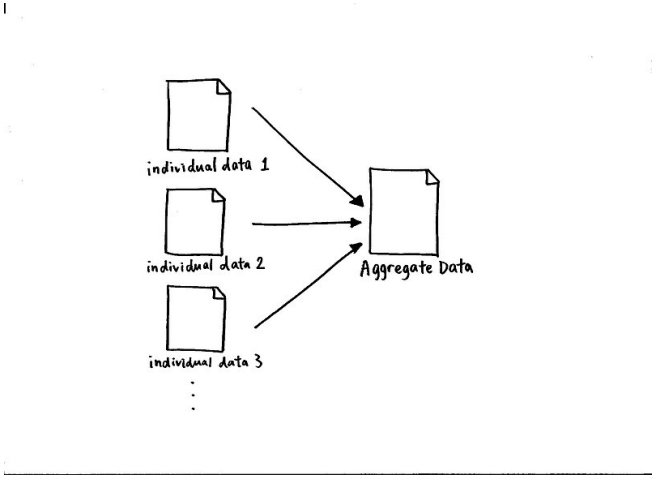
Group

- Grouped data are data formed by aggregating individual observations of a variable into groups, so that a frequency distribution of these groups serves as a convenient means of summarizing or analyzing the data.



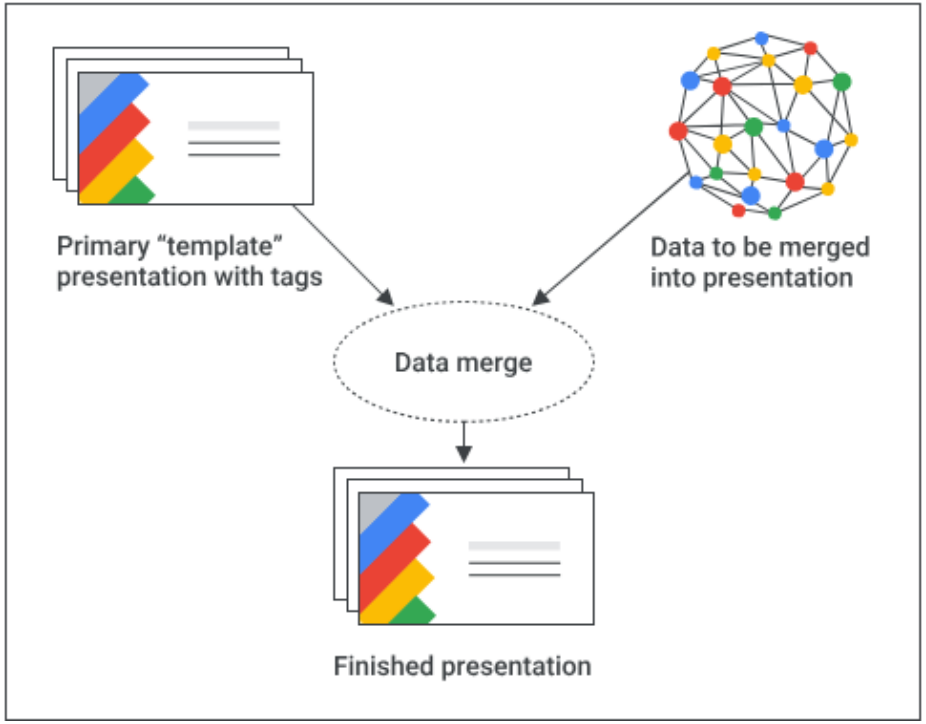
Aggregate

- Data aggregation is the process where raw data is gathered and expressed in a summary form for statistical analysis.
- Aggregate data is high-level data which is acquired by combining individual-level data.
- AVERAGE, SUM, PRODUCT, COUNT, COUNTA, MAX, or MIN



Merge

- Data merging is the process of combining two or more similar records into a single one.



Pivot

- Data pivoting enables you to rearrange the columns and rows in a report so you can view data from different perspectives.

Country	Year	Profit (USD)
USA	2020	495875
USA	2021	459875
France	2020	145685
France	2021	201457
Germany	2020	178563
Germany	2021	165478

Pivot →

Country	2020	2021
USA	495875	459875
France	145685	201457
Germany	178563	165478

Unpivot →

Country	Year	Profit (USD)
USA	2020	495875
USA	2021	459875
France	2020	145685
France	2021	201457
Germany	2020	178563
Germany	2021	165478

Windowing

- Window functions are sometimes used in the field of statistical analysis to restrict the set of data being analyzed to a range near a given point.


рахмат
danke
謝謝
ngiyabonga
شكراً جزيلاً
spas
teşekkür ederim
Баярлалаа
mersi
kia ora
barka
welalin
tack
mahalo
tapadh leat
спасибо
faafetai lava
vinaka
blagodaram
dank je
misaotra
matondo
paldies
grazzi
спасибі
kiitos
dankie
nanni
nandri
dhanyavad
thank you
gracias
хвала
asante
manana
mauruuru
köszönöm
açiü
dankon
akun
hvala
gràcie
bayarlalaa
bedankt
enkosi
obrigada
murakoze
tenki
djiere dieuf
tau
mochchakkeram
дякую
mamnun
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
chokrane
sobodi
dĕkuji
sukriya
kop khun krap
obrigado
sagolun
arigatô
takk
dakujem
trugarez
mèsi
didi madloba
ありがとう
tanemirt
grazie
najis tuke
terima kasih
rahmet
ευχαριστώ
diolch
dhanyavadagalu
shukriya
mercé
мерси
kam sah hamnida
rahmat
তোমাকে ধন্যবাদ
감사합니다
xièxie
merci