Data Analysis Fundamentals

Lecture 2: Understanding of Data

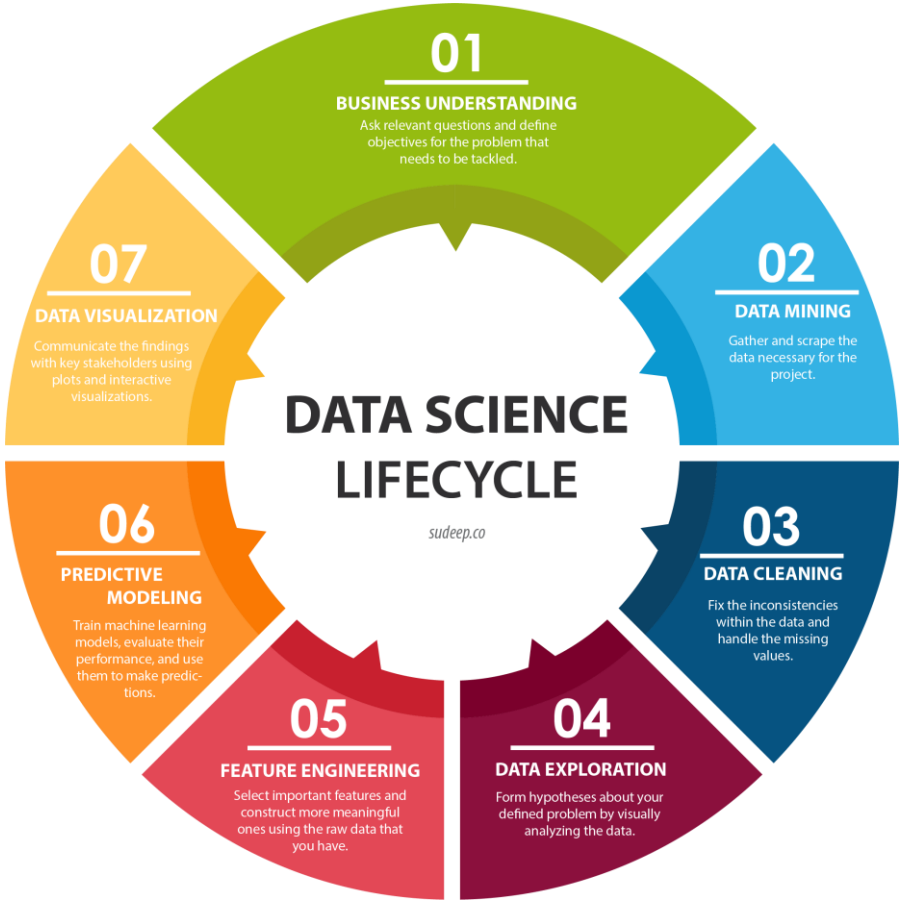
Vikas Kumar
Indian Institute of Technology Bombay

05th September 2023

Outline

- Understanding phases of a typical Data Analytics Project
- Understanding Data
- Derived Facts/Dimensions
- Building dimensions from Facts (Binning)
- Granularity of Data

Data Science/Analysis Pipeline



Understanding Data

Qualitative Data
(Categorical)

- Gender
- Religion
- Marital status
- Native language
- Social class
- Qualifications
- Type of instruction
- Method of treatment
- Type of teaching approach
- Problem-solving strategy used

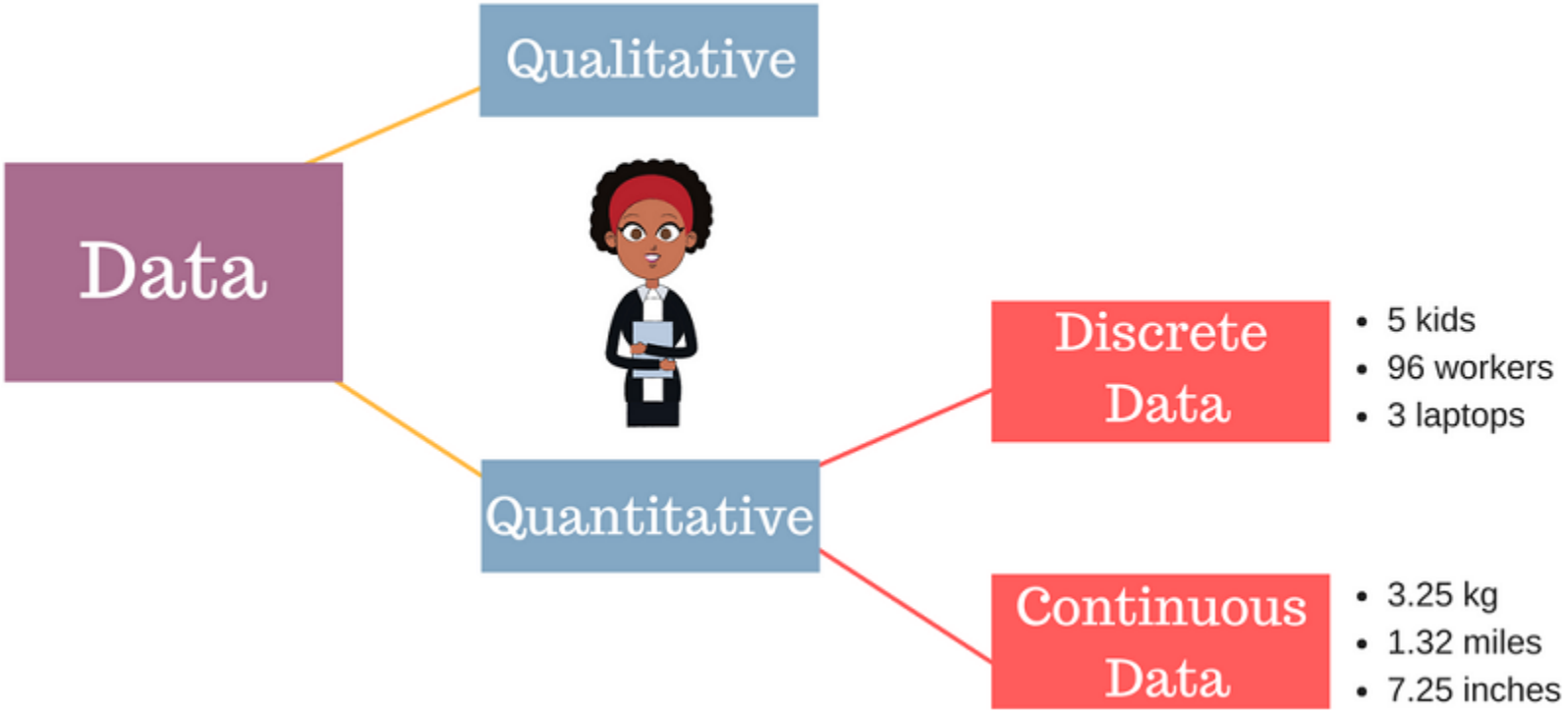
Quantitative Data
(Numerical)

- Age
- Height
- Weight
- Income
- University size
- Group size
- Self-efficacy test score
- Percent of lecture attended
- Clinical skills performed
- Number of errors

Quantitative Data/Facts/Metrics/Measures/KPIs

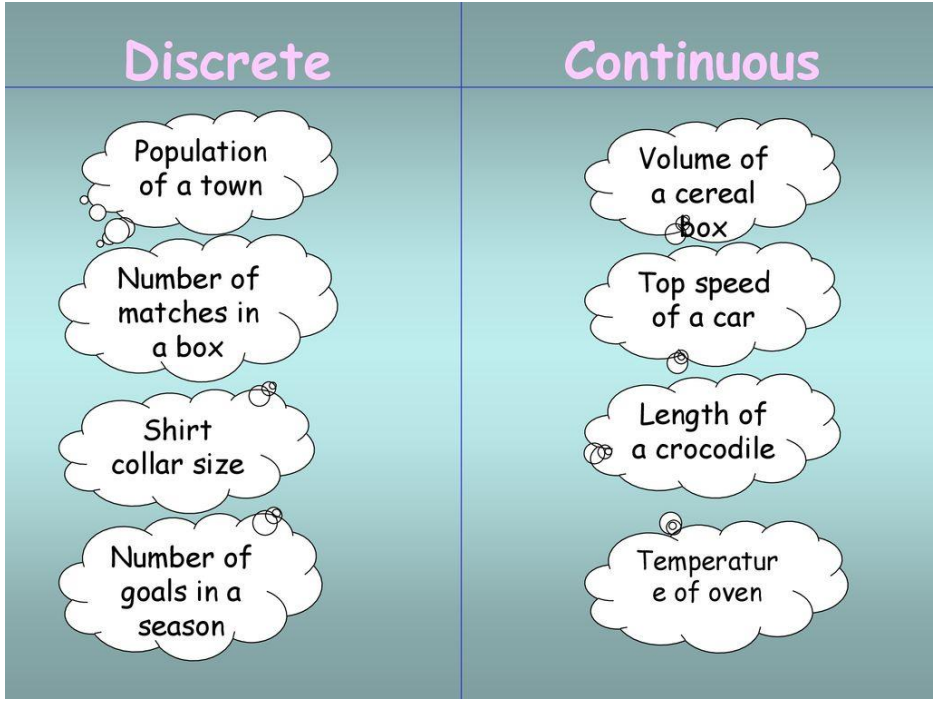
KPIs	Metrics
All KPIs are Metrics	All Metrics are not KPIs
KPIs give a holistic view of the performance of different functions in your organization	Metrics give you a picture of how different individual activities rolled out within the functions are progressing
KPIs tell you where exactly your teams stand with respect to the overall business goals	Individual Metrics do not give any insights on their own
Examples: Pre-sales KPIs, Email Marketing KPIs, Customer Success KPIs	Examples: Open Rate, Conversations in the last 2 weeks, Deals lost last quarter

Continuous vs Discrete data



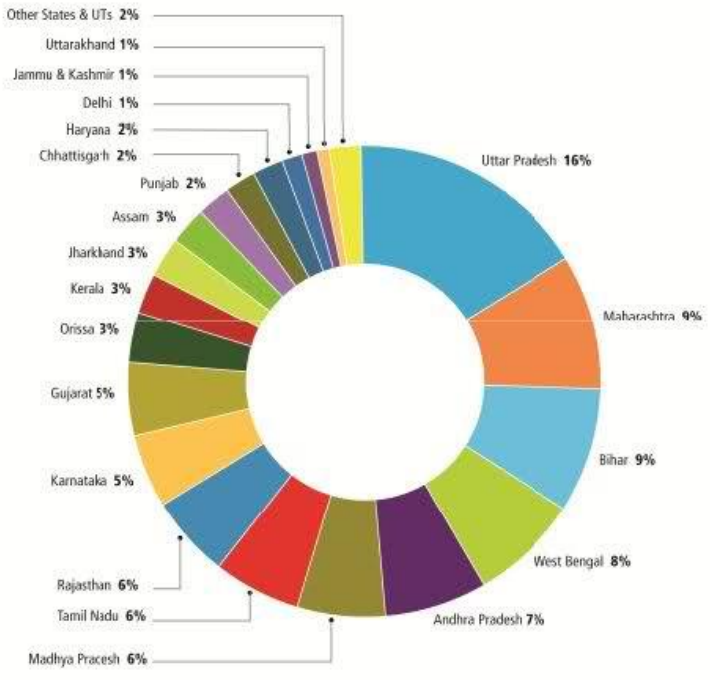
Continuous vs Discrete data

- Discrete data is a finite value that can be counted.
- Continuous data has an infinite number of possible values that can be measured.



Qualitative Data/Dimensions/Perspectives/Slicers

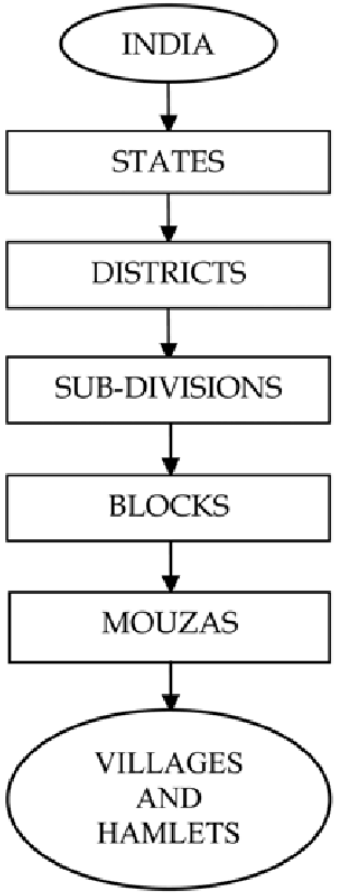
Qualitative Data/Dimensions/Perspectives/Slicers

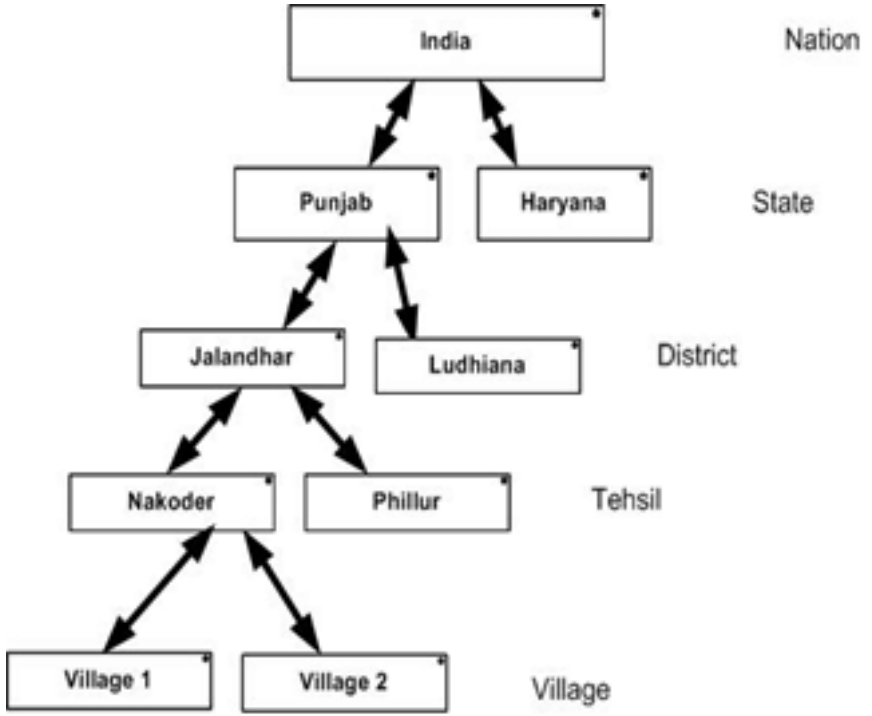


Share of different States in India's population

Our Census, Our Future

Hierarchical Dimensions





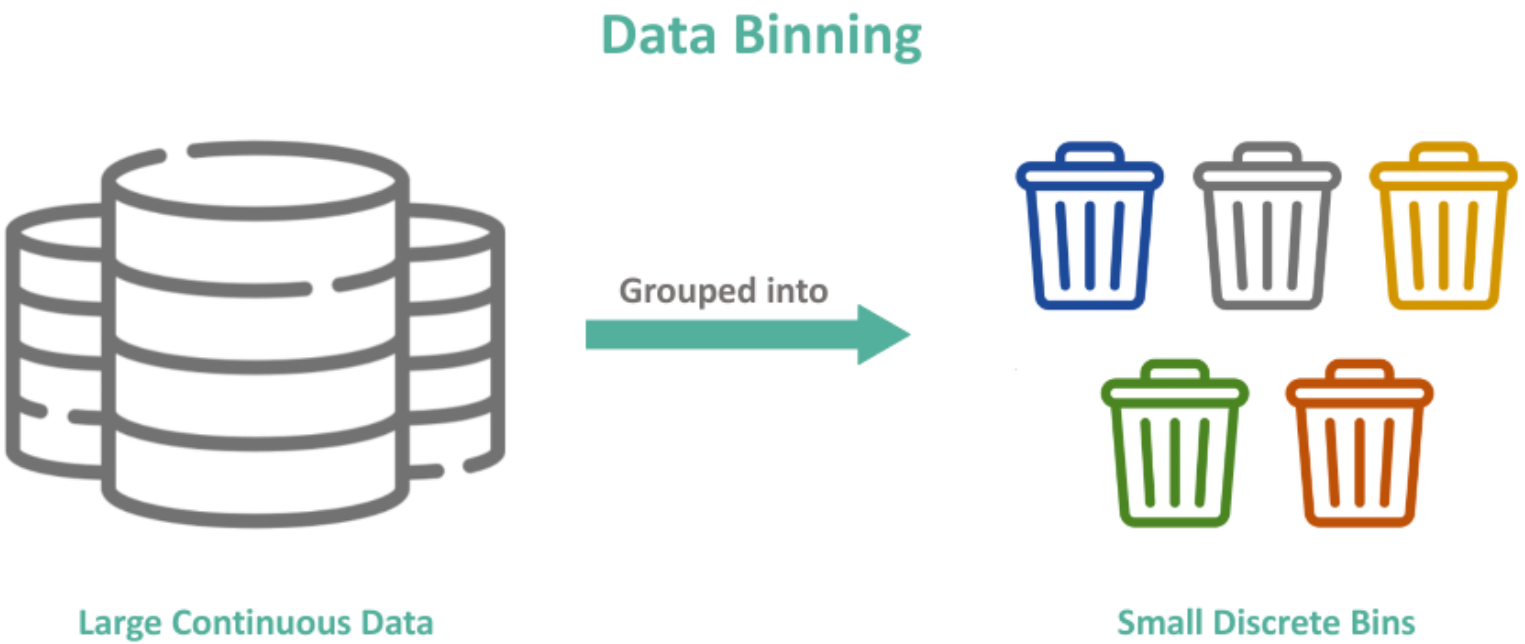
Derived Facts/Dimensions

- In data warehousing, facts and dimensions are standard terms.
- They inform us about things like the number of resources used for a particular task.
- They both store the exact measure of resources and details about the resource and task.
- A fact in data warehousing describes quantitative transactional data like measurements, metrics, or values ready for analysis.
- Dimensions are companions to facts and are attributes of facts like the date of a sale.

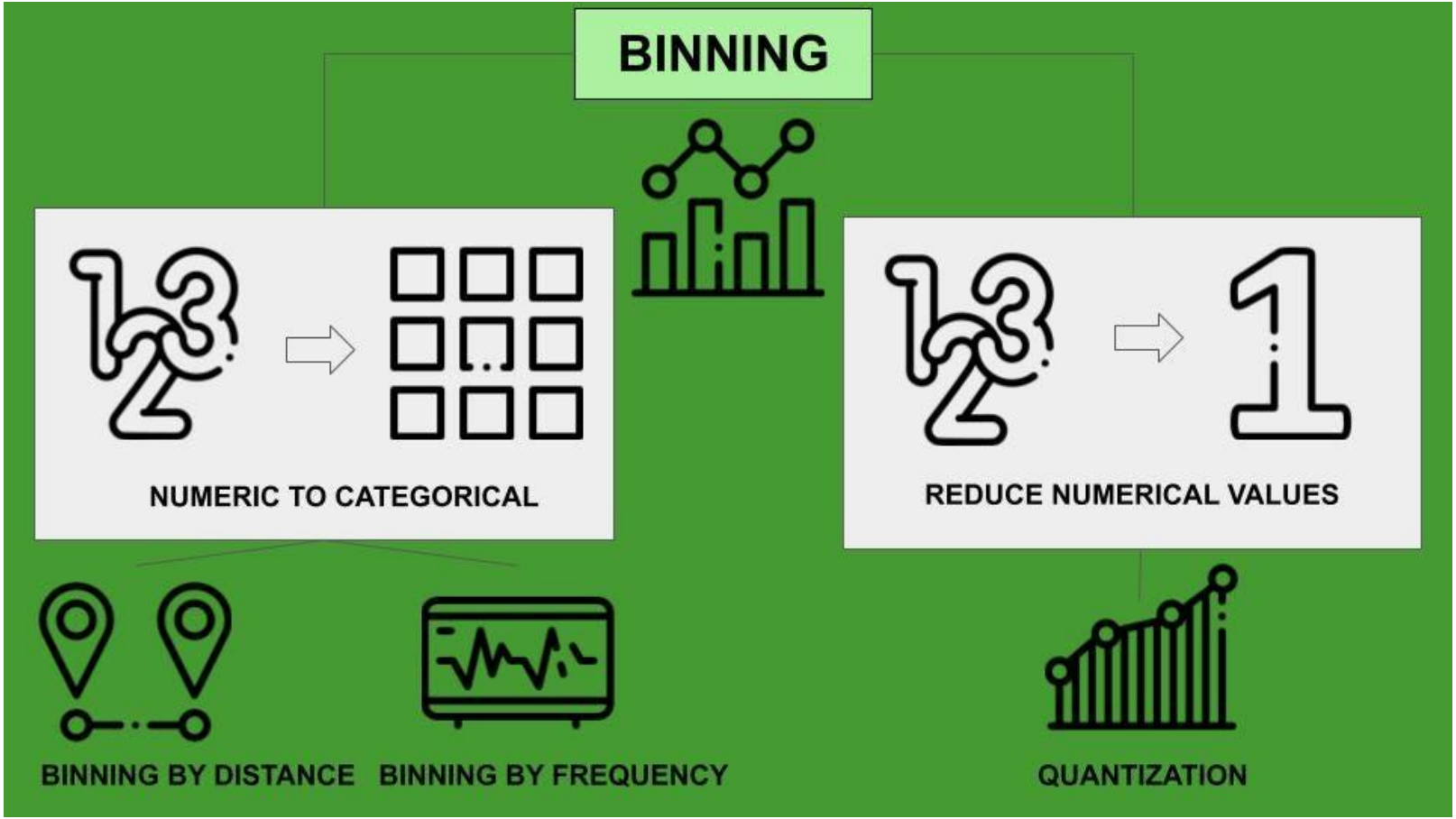
Building Dimensions from Facts (Binning)

- Data binning, also called discrete binning or bucketing, is a data pre-processing technique that reduces the effects of minor observation errors.
- The original data values are divided into small intervals known as bins, and then they are replaced by a general value calculated for that bin.
- Data binning is a way to group numbers of more or less continuous values into a smaller number of "bins".

Building Dimensions from Facts (Binning)



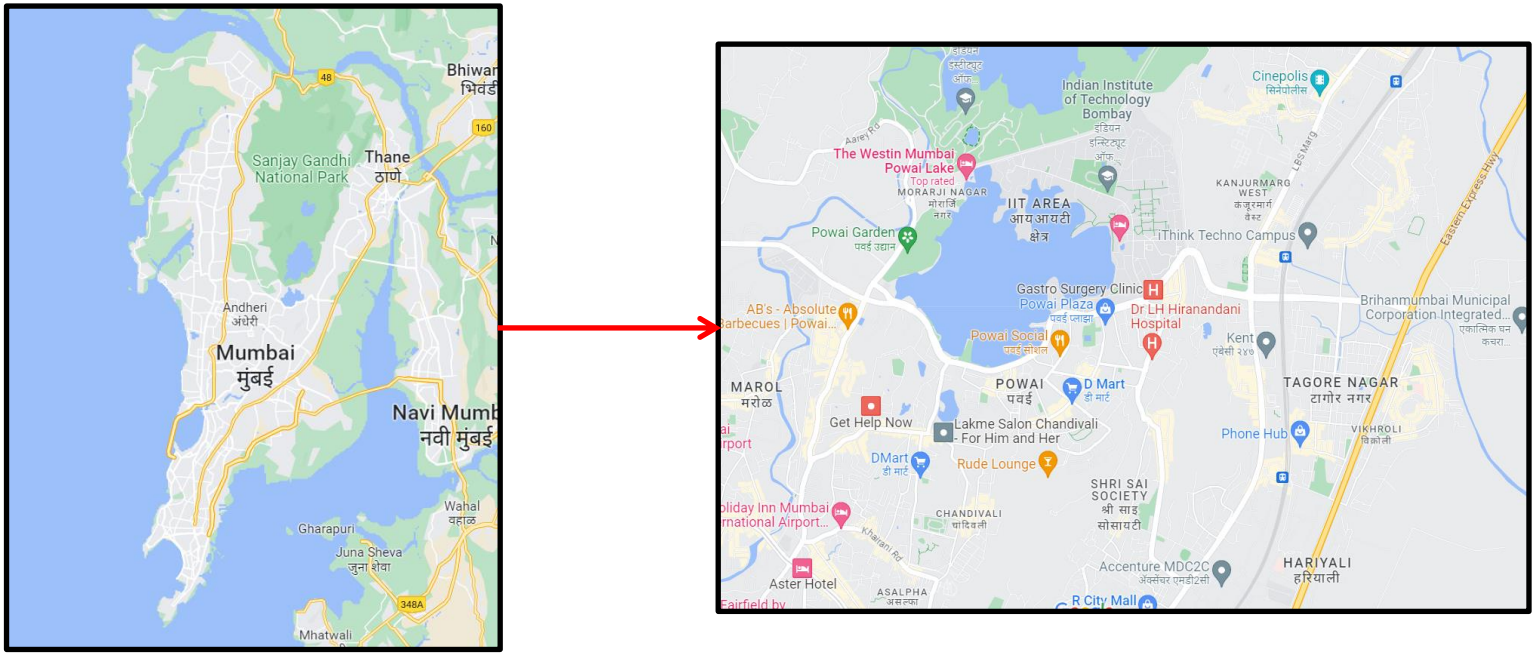
Building Dimensions from Facts (Binning)



Granularity of Data

- Data granularity measures the level of detail in a data structure.

- In time-series data, for example, the granularity of measurement might be based on intervals of years, months, weeks, days, or hours.



Identifying/Constructing unique keys

Features	**Primary keys**	**Unique keys**
Number of keys	One primary key in a parent table	One or more than one, in parent or child tables
Values	Must have a value, cannot be NULL	Can be a NULL value
Use	Identify every item in a table	Identify items in a table when they cannot have duplicate values
Ease	Cannot be removed, difficult to change	Can be removed or changed easily
Indexes	Clustered index	Non-clustered index

Identifying/Constructing unique keys

Student

Roll_no	Name	Class	Phone_no	Registration_no
1	Andrew	5	9854672256	895
2	Andrew	6	9955512456	564
3	Augosto	5		567

Primary key

Unique key

Unique key


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