

SNS COLLEGE OF TECHNOLOGY

Coimbatore-35.

An Autonomous Institution

COURSE NAME : DATA ANALYTICS

II YEAR/ IV SEMESTER

UNIT – II Getting Insights from Data

Topic: Infographics & Word Clouds

Dr.K.Sangeetha

HoD

Department of Computer Science and Engineering

Infographics

- Transform data into an image
- info-graphic approach is subjective
- produced manually and is customized for a Particular data set.

Data visualization, on the other hand, is objective, automatically produced and can be applied to many data sets

HEALTHY EATING PLATE

HEALTHY OILS

Use healthy oils (like olive and canola oil) for cooking, on salad, and at the table. Limit butter. Avoid trans fat.

The more veggies – and the greater the variety – the better. Potatoes and French fries don't count.

Eat plenty of fruits of all colors.

STAY ACTIVE!

© Harvard University

WATER

Drink water, tea, or coffee (with little or no sugar). Limit milk/dairy (1-2 servings/day) and juice (1 small glass/day). Avoid sugary drinks.

WHOLE GRAINS

Eat a variety of whole grains (like whole-wheat bread, whole-grain pasta, and brown rice). Limit refined grains (like white rice and white bread).

HEALTHY PROTEIN

Choose fish, poultry, beans, and nuts; limit red meat and cheese; avoid bacon, cold cuts, and other processed meats.

Harvard T.H. Chan School of Public Health
The Nutrition Source
www.hsph.harvard.edu/nutritionsource

Harvard Medical School
Harvard Health Publications
www.health.harvard.edu

THE JOURNEY OF BIRYANI IN INDIA

BIRYANI
Originated in Persia and was brought to South Asia by Muslim Merchants to Mughal Empire

KASHMIRI BIRYANI
Addition of asafoetida

MUGHLAI BIRYANI
Meat cooked in rich spices

LUCKNOWI BIRYANI
Milder taste

KAMPURI BIRYANI
Mix of chicken & veggies

KOLKATA BIRYANI
Known for use of potatoes

HYDERABADI BIRYANI
Extremely Spicy

BOMBAY BIRYANI
Typically sweeter

PILAF BIRYANI
Subtly spiced

GOAN FISH BIRYANI
Heavily spiced & aromatic

BENGALURU BIRYANI
Cooked with coriander & mint

DINDUGAL THALAPAKATTI BIRYANI
Unique taste & rich spicy aroma

MALABAR BIRYANI
Light & aromatic

Spicy

The word Biryani is derived from the Persian word Birian, which means 'fried before cooking' and Birinj, the Persian word for rice.

the yellow chilli
REDEFINING INDIAN FOOD

Word Cloud

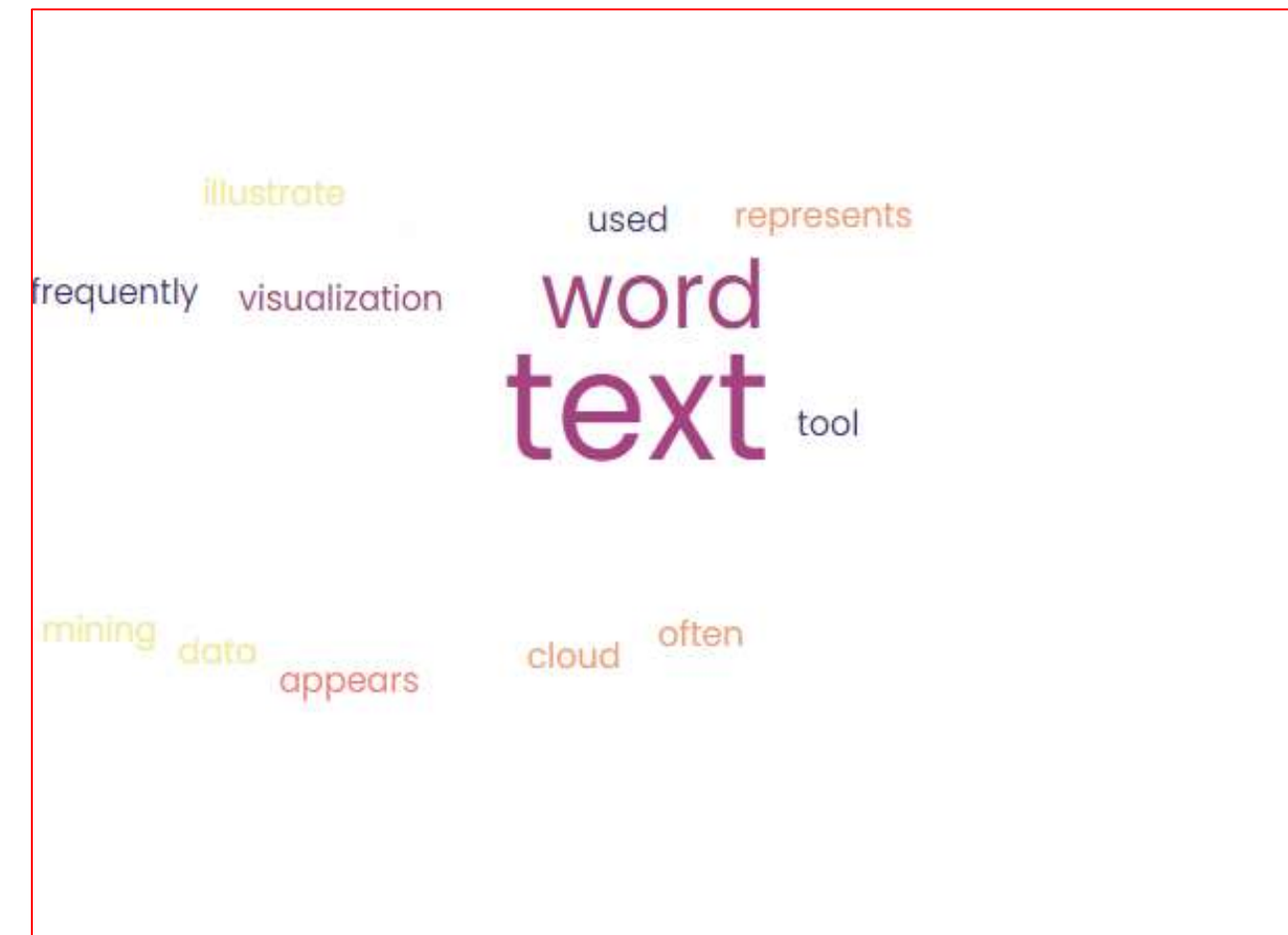
A visualization tool frequently used in text mining to illustrate text data is the word cloud which Represents how often each word appears in a given text

The higher the frequency of a word in the text, the larger its size in a word cloud.

The screenshot shows a web interface for generating word clouds. It has two main sections: 'Paste/Type Text' and 'Options'. The 'Word List' modal is open, displaying a table with 14 words and their frequencies. The table is as follows:

Frequency	Word	Checkmark
3	text	<input checked="" type="checkbox"/>
2	word	<input checked="" type="checkbox"/>
1	visualization	<input checked="" type="checkbox"/>
1	tool	<input checked="" type="checkbox"/>
1	frequently	<input checked="" type="checkbox"/>
1	used	<input checked="" type="checkbox"/>
1	mining	<input checked="" type="checkbox"/>

At the bottom of the modal, there are three checked options: 'Remove Stop Words', 'Remove Numbers', and 'Remove Special Characters'. The background interface includes a 'Paste/Type' button and a 'WORDCLOUD' button.



Data Quality & Preprocessing

Data Quality

Quality of the models, charts and studies in data analytics depends on the quality of the data being used

The application domain, human error, the integration of different data sets and the methodology used to collect data can generate data sets that are noisy, inconsistent, or contain duplicate records

Data quality is important and can be affected by internal and external factors

Internal factors can be linked to the measurement process and the collection of information through the attributes chosen.

External factors are related to faults in the data collection process, and can involve the absence of values for some attributes and the voluntary or involuntary addition of errors to others.

1)Missing Values

- attributes values only recorded some time after the start of data collection, so that early records do not have a value
- the value of an attribute being unknown at time of collection
- distraction, misunderstanding or refusal at time of collection
- attribute not required for particular objects
- non-existence of a value
- fault in the data collection device
- cost or difficulty of assigning a class label to an object in classification problems

Solutions

- Ignore missing values
- Remove objects
- Make estimates

Example 4.1 As an example of how to deal with missing values, let us consider the data set in Table 4.1. Suppose that, due to a data transmission problem, part of our contact data sent to a colleague was missing. Table 4.1 shows how missing values in the data set can be filled, using the mode for qualitative values

Table 4.1 Filling of missing values.

Data with missing values				Data without missing values			
Food	Age	Distance	Company	Food	Age	Distance	Company
Chinese	51	Close	Good	Chinese	51	Close	Good
			Good	Chinese	53	Close	Good
Italian	82		Good	Italian	82	Close	Good
Burgers	23	Far	Bad	Burgers	23	Far	Bad
Chinese	46		Good	Chinese	46	Close	Good
Chinese			Bad	Chinese	31	Far	Bad
Burgers		Very close	Good	Burgers	53	Very far	Good
Chinese	38	Close	Bad	Chinese	38	Close	Bad
Italian	31	Far	Good	Italian	31	Far	Good

- Redundant data
- Inconsistent Data