

Name - Shubham Khot | Rollno- 3025  
Class - TE (IT) T2 | sub - DS BDB  
Assi - 18/20 Marks.

(9)

→ Theory Assignment - 3 →

Date: 21/4/25

Q1. What are challenges in Big Data visualizations? Why it is needed?

→ Some challenges in Big Data visualization -

- ① Visual noise - In which user struggle to distinguish b/w different objects in dataset due to sheer volume of data being displayed
- ② Information loss - This issue occurs when efforts to reduce visual clutter lead to omission of important data.
- ③ Additionally, sheer volume of big data can overwhelm user, making it difficult to discern key insights from vast datasets.
- ④ Limited customization tools - In some tools, which can hinder creation of tailored visual representations that meet specific user needs.
- ⑤ Prof - Because it helps in interpreting & communicating complex data effectively. It enables user to draw conclusions quickly & easily.
- ⑥ Efforts to improve it include adopting scalable cloud storage solutions, implement deduplication, data integration.



Date: \_\_\_\_\_

Q. 2. Explain tools - 1) candelab

2) Google chart API 3) D3.js

=)

1) Candelab - opensource visualization tool,

- It is visualization library designed for creating high-quality, scalable & reusable visualization.

- It is built on top of other libraries like vega, D3.js, WebGL.

- Candelab focuses on scientific, statistical & data-driven visualization.

- consistent & easy to use interface supports interactive charts.

- cross-platform web app & notebooks.

2) D3.js - [Data-Driven Documents]

- Javascript library used to create dynamic & interactive highly customized data visualizations in web browser.

- D3 binds Data into HTML, SVG & CSS elements, making it easy to manipulate & animate.

- Highly extensible & modular.

- Allows creation of completely customized visualizations.

- Supported by large developer community.

- More powerful & robust steep learning curve.

### Q) Google Chart API

- It's a free tool provided by Google for creating simple & interactive charts.
- It uses Javascript to embed charts into web pages easily.
- It can create different types of charts like line, bar, pie, bubble, geo charts & more.
- Simple to use, responsive.
- Provides interactive features - tooltips, zooming, filtering.
- Works seamlessly with Google Sheets to fetch live data.
- Free of cost, no complex coding.

### Q. Explain Michael Porter's Analysis Model.

- $\Rightarrow$  - It's a known as Porter's Five force model.
- It is a framework used to analyze competitive forces within industry.
  - It helps organization to understand strength of their market position & profitability of their industry.

#### Five Forces of Porter's Model

- 1) Competitive Rivalry - Refers to intensity of competition among existing competitors in a market.
- High competition limits profitability of companies to lower prices or increase money on marketing.



- (2) Threat of New Entrants - Refers to risk of new companies entering market.
- It lowers profitability.
  - Barrier to entry -
    - High startup cost
    - Government regulations
    - Economic scale
    - Strong brand of existing players

(3) Bargaining power of Suppliers -

- Suppliers can influence cost of quality of inputs.
- If there are few suppliers and if a supplier offers a unique product, their bargaining power is high.
- High supplier power can increase input cost.
- Reduces profit margin.

(4) Bargaining power of Buyers -

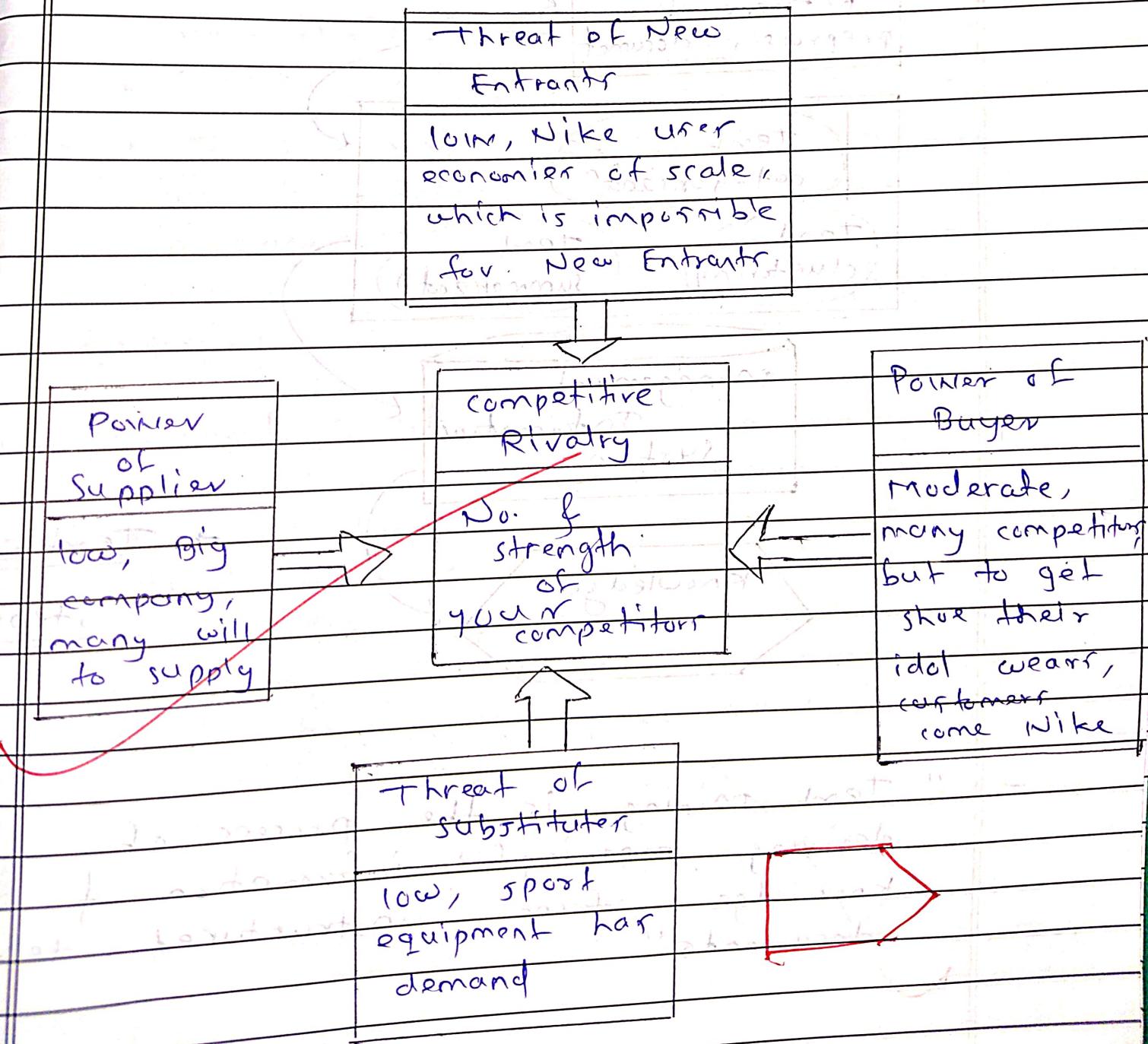
- Customers can demand lower price, higher quality, more services.
- It causes companies to compete harder to retain them.
- Price sensitivity, Volume of purchases can affect this factor.

(5) Threat of Substitute Products - Refers to risk that customers will switch to different products.

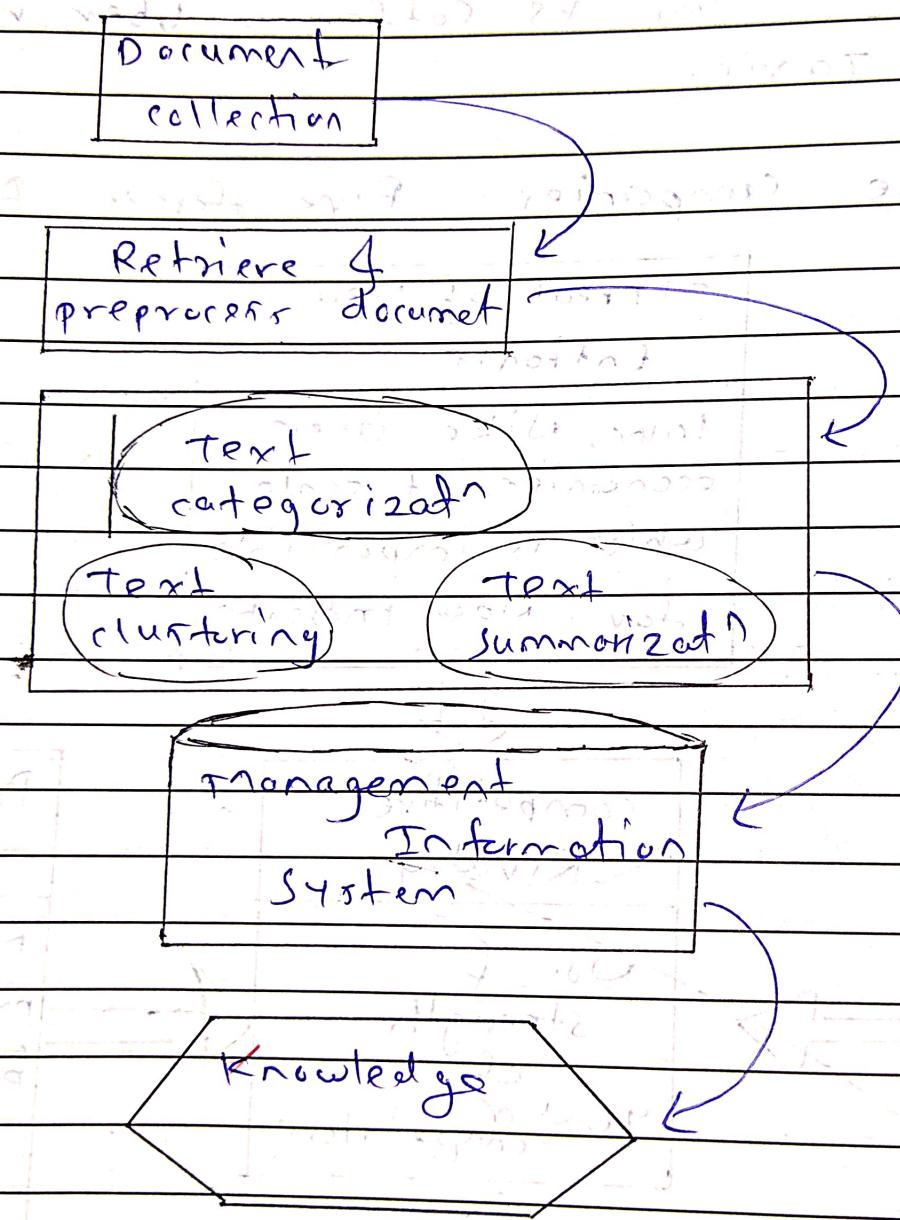


- More substitutes available, greater pressure on companies for innovation & reduce price.
- Ex - Tea vs Coffee, Uber vs Local Taxis.

### Ex Nike Company Five Forces Analysis -



Q.4. What is text-mining? Explain with the help of diagram of steps.



~~"Text mining is the process of deriving meaningful information & knowledge from unstructured text documents".~~

- Techniques used - NLP, Info Retrieval, machine Learning, & Data mining to analyze, categorize, summarize & extract patterns from text

### Why Text mining?

- 80% of all data generated today is unstructured text. (emails, reports, reviews)
- Traditional methods can't directly work.
- Text mining helps organizations gain insights, detect trends, predict customer behaviour.

### Steps

#### (1) Document collection

- Raw text is collected.
- Sources can be news articles, emails, social media, customers review, research papers.
- Goal - Gather large text data for mining.

#### (2) Retrieve & Preprocess Document

- Collected documents are retrieved & cleaned to make them suitable for analysis.

- Common preprocessing techniques
  - Tokenization
  - Stop word Removal
  - Stemming
  - Lowercasing.

### (3) Text Categorization, Clustering, & Summarization

- After preprocessing, 3 main operations are performed.
- 1) Categorizing documents
    - ex - emails into "spam" or "ham".
  - 2) Reclustering similar documents
    - ex - Grouping news about similar events
  - 3) Summarization - shortening documents
    - ex - long research papers → bullet points.

### (4) Management Info. System

- All text is stored in MITS.

- It allows querying, analyzing, visualizing the precessed data easily.

### (5) Knowledge Extraction

- Final step: not yet done.

- Useful trends, patterns, & decisions are extracted.

knowledge helps in

- BI

- Decision making

- Predictive Analysis.