

# Assignment - 2

## Data as an Asset

---

Harsh Siddhapura  
1230169813

IFT511 - Big Data Analysis  
Prof. Asmaa Elbadrawy

18th January, 2024

## PROBLEM - 1

1. To recommend videos for a user, YouTube would need various types of data related to users, videos, and their interactions such as:

- Likes
- Channels Subscribed
- Watch History
- Comments
- Video Category
- Views
- Hours Spent
- Location
- Timezone
- Advertisement
- Engagement
- Interactions
- Live Events
- Sharing
- Active Hours

The data could be organized into several tables:

User ID	Username	Subscribed Channels	Watch History
1	Nancy123	Channel1, Channel2	Video1, Video2
2	User2	Channel2, Channel3	Video3, Video4
...	...	...	...

Figure-1: User Table

Video ID	Title	Category	Views
101	Video1 Title	Music	500,000
102	Video2 Title	Gaming	1,200,000
...	...	...	...

Figure-2: Video Table

Interaction ID	User ID	Video ID	Interaction Type
201	1	101	Watch
202	2	102	Like
...	...	...	...

Figure-3: User Interaction Table

2. The storage of user data for generating recommendations raises important data privacy concerns. Users' watch history, subscriptions, and interactions provide insights into their preferences and behavior. While data storage is crucial for enhancing user experience and providing personalized recommendations, it also poses risks to user privacy.

From a data privacy perspective, it is essential for platforms like YouTube to prioritize user consent, transparency, and data security. Users should have control over what data is collected and how it is used. Clear and user-friendly privacy policies, robust security measures, and options for users to manage their privacy settings are crucial.

However, there's often a trade-off between personalized recommendations and user privacy. Striking the right balance involves implementing privacy-enhancing technologies, anonymizing data where possible, and ensuring that user data is used responsibly and ethically.

It's important for platforms to be transparent about their data practices and to continually assess and improve privacy safeguards. User trust is fundamental, and respecting user privacy is key to maintaining that trust.

## PROBLEM - 2

1. Conducting a long-term study to evaluate the side effects of a drug involves collecting comprehensive data on patients' health, medical history, and drug usage over an extended period. The data needed could include:
  - Demographic Information: Age, gender, ethnicity, and other relevant demographic details.
  - Baseline Health Status: Existing health conditions, family medical history, lifestyle factors, and baseline measurements.
  - Drug Usage Information: Dosage, frequency, duration of drug usage, adherence to prescribed regimen.
  - Health Monitoring: Regular health check-ups, lab tests, and imaging studies to monitor physiological parameters.
  - Adverse Events: Any reported adverse events or side effects during the study period.
  - Medical Interventions: Any medical interventions or changes in drug regimens during the study.
  - Follow-Up Surveys: Periodic surveys to assess patients' subjective experiences and well-being.
  - Lifestyle Factors: Changes in lifestyle, dietary habits, exercise routines, and other relevant factors.
  - Health Outcomes: Monitoring of specific health outcomes (e.g., cardiovascular events, organ function).
2. A table could be organized to represent each patient's data over time.
  - Each row would represent a unique patient
  - Columns would represent various attributes tracked over the study period

Patient ID	Demographics	Baseline Health	Drug Usage	Health Monitoring	Adverse Events	Interventions	Follow-Up Surveys	Lifestyle Factors	Health Outcomes
1	...	...	...	...	...	...	...	...	...
2	...	...	...	...	...	...	...	...	...
....	...	...	...	...	...	...	...	...	...
10000	...	...	...	...	...	...	...	...	...

3. It may be feasible to do this study because:

- Benefits:
  - Understanding the long-term side effects of a drug is crucial for patient safety and regulatory compliance.
  - Positive outcomes could enhance the drug's reputation and marketability.
  - May lead to improved patient care and overall public health.
- Challenges:
  - High costs associated with the long-term study, including data collection, monitoring, and analysis.
  - Unpredictable outcomes—there is no guarantee that significant side effects will be discovered.
  - Competitive and regulatory pressures may impact the commercial viability of the study.
- Considerations:
  - The feasibility of such a study depends on the specific drug, its market potential, and regulatory requirements.
  - Pharmaceutical companies weigh potential benefits against costs and regulatory obligations.
  - Collaboration with regulatory agencies and academic institutions may provide funding and expertise.

Despite the challenges and costs, the ethical and legal responsibilities of pharmaceutical companies often necessitate rigorous testing, especially for long-term drug safety.

Balancing business interests with ethical considerations is critical for sustaining trust in the pharmaceutical industry.

In summary, the feasibility of the study depends on a complex interplay of ethical, regulatory, and business considerations. Careful planning, collaboration, and ethical considerations are essential for the success of such long-term studies.