Task 1
Pan Zhangyu P136922

1.

Example of raw data: https://www.iqair.cn/cn/malaysia/kuala-lumpur

Data: The raw data consists of numbers collected by air quality monitors in the city, such as concentrations of pollutants like PM2.5 and PM10, nitrogen dioxide, and ozone. These data are just numbers without any processing or interpretation.

Information: When we process and interpret this data, it becomes information. For example, we can compare these data with national or international air quality standards to determine if the city's air quality meets safety criteria. These comparisons and analyses provide information about the air quality in the city.

Knowledge: Based on this information, we can start to form knowledge. For instance, we can analyze historical data to identify which factors contribute to worsening air quality and how measures can be taken to improve it. This deeper analysis and understanding constitute knowledge, allowing us to comprehend the nature of air quality issues.

Wisdom: Ultimately, when we translate this knowledge into action, we reach the top of the pyramid — wisdom. By implementing appropriate policies, actions, and measures, we can improve the city's air quality and ultimately enhance the quality of life for its residents. This wisdom is based on a deep understanding and analysis of data, information, and knowledge, as well as rational decision-making regarding actions.

Therefore, the process from raw data to wisdom is incremental, involving steps of data processing, information interpretation, knowledge formation, and wisdom application, as depicted in the DIKW pyramid.

2.

With the proliferation of digital technologies, unstructured data is generated from various sources such as social media, sensors, mobile devices, and IoT devices. These data include text, images, videos, audio, etc., which are closer to real-world scenes

and experiences. Unstructured data often contains rich and valuable information that may not be captured through structured data alone. For example, text data from social media posts or customer reviews can provide insights into customer sentiment, preferences, and behaviors that are critical for businesses to understand and respond effectively. Organizations that effectively leverage unstructured data can gain a competitive advantage by making data-driven decisions, improving customer experience, optimizing operations, and discovering new business opportunities. Unstructured data provides a more complete and integrated view of the business environment, enabling organizations to stay ahead of the curve.

3.

NETFLIX collects user data to understand what each customer wants to see, and based on this data can provide customers with content they are interested in. Customers will also be more satisfied because they can see what they want without searching. Customer retention rates will also increase. Credit card companies will collect data such as the time and place of card swiping, which will help prevent fraud. For example, it is impossible for a person to swipe his card in two places thousands of kilometers apart in a short period of time.

4.

a. Bank:

Risk management: Banks can use big data to analyze customers' historical transaction data, credit scores, social media information, etc. to identify potential credit risks and take corresponding risk management measures.

Anti-fraud: By analyzing large amounts of transaction data and behavioral patterns, banks can detect and prevent fraud and protect customer assets and bank interests.

b. Correspondence:

Customer insights: Communication companies can use big data to analyze customers’ call records, usage habits and location data to understand customer needs and preferences and provide personalized services and recommendations.

Network optimization: By analyzing network traffic, signal coverage and device connection data, communications companies can optimize network configuration and resource allocation, improving network performance and user experience.

c. Medical insurance:

Personalized treatment: Medical institutions can use big data to analyze patients' genomic data, medical records, and lifestyle information to provide patients with personalized diagnosis and treatment plans.

Disease surveillance: By monitoring large-scale health data and epidemiological information, medical institutions can promptly identify and control the spread of diseases and protect public health.

d. Media:

Content recommendation: Media companies can use big data to analyze users' browsing history, click behavior, and social media interactions to recommend personalized content and ads to users.

Audience insights: By analyzing viewers' viewing habits, preferences, and comment feedback, media companies can understand their audiences' interests and needs and develop more effective content strategies.

e. Advertise:

Target positioning: Advertising companies can use big data to analyze users' search history, purchasing behavior and social media information to accurately locate target audiences and provide relevant advertising content.

Advertising effectiveness evaluation: By analyzing data such as ad exposure, click-through rates, and conversion rates, advertising companies can evaluate the effectiveness of advertising campaigns and optimize advertising strategies.

f. Manufacture:

Predictive maintenance: Manufacturing companies can use big data to analyze equipment sensor data and production line operations, predict equipment failures and maintenance needs, and reduce production interruptions and repair costs.

Quality control: By analyzing sensor data and quality inspection records during the production process, manufacturing companies can monitor product quality in real

time and discover and correct production problems in a timely manner.

g. Transportation:

Intelligent logistics: Transportation companies can use big data to analyze cargo transportation data, traffic conditions, weather information, etc., optimize logistics routes and transportation plans, and improve transportation efficiency and punctuality.

Customer experience: By analyzing passengers' travel habits, itinerary preferences and complaint feedback, transportation companies can improve service quality and enhance customer experience.

h. Retail:

Precision marketing: Retailers can use big data to analyze customer purchase history, behavioral data and social media activities to provide customers with personalized promotions and coupons.

Inventory management: By analyzing sales data, seasonal demand and supply chain information, retailers can optimize inventory management and reduce overstocking and out-of-stock situations.

5.

Aspect	Structured Data	Unstructured Data	Semi-Structured Data
Definition	Data that conforms to a pre-defined schema or format. Examples include relational databases, spreadsheets.	Data that lacks a pre-defined structure or format. Examples include text documents, images, videos.	Data that does not fit neatly into either structured or unstructured categories. Examples: NoSQL.
Sources	Generated from organized systems and applications such as sensors, Weblogs, all data that humans enter into computers.	Generated from various sources such as social media, scientific data, radar data.	Generated from sources that have some structure but may also contain unstructured elements, such as log files, web server logs.
Feature	Accounting for about 20% of the total data. Most used in computer related activities.	The remaining data created accounts for about 80% of the total. Until recently, there wasn't much that could be done except storage and manual analysis.	Much semi-structured data appears to be unstructured at first glance.