1. Discuss whether each of the following activities is a data mining task:

Note – Data Mining: is the process of automatically discovering useful information in large data repositories. Data mining techniques are deployed to scour large datasets to find novel and useful patterns that might remain unknown. They also provide the ability to predict the outcome of future observations. (Chapter 1, Page 4)

1. Dividing the customers of a company according to their gender: This activity is not considered data mining because it can be seen as a simple data classification task and does not involve discovering patterns or relationships in the data. I would argue this is a data organization task.
2. Dividing the customers of a company according to their profitability: This activity is not necessarily data mining on its own, but it could be depending on if one is assessing multiple factors to assess customer profitability.
3. Computing the total sales of a company: This activity is not data mining. One may interpret this as a simple aggregation that falls under exploratory data analysis.
4. Sorting a student database based on student identification numbers: This is not a data mining task. This is a data organization task.
5. Predicting the future stock price of a company using historical records: I would interpret this as a data mining task. Analyzing historical records and identifying patterns, trends, and relationships can be applied to predict future stock prices.
6. Monitoring the heart rate of a patient for abnormalities: This is not a data mining task because one is just observing heart rates. However, one can argue that this can lead to a formalized data mining process.
7. Monitoring seismic waves for earthquake activities: This is not data mining. It is a data analysis task that does not involve discovering patterns or relationships in the data.
8. Extracting the frequencies of a sound wave: Extracting the frequencies of a sound wave is not data mining. It is a signal processing task that involves analyzing the properties of a sound wave.
9. As an analyst for an internet research company, data mining techniques can be extremely helpful for the overall performance and serve as a differentiator of the search engine. This is how a company can benefit from the following data mining techniques:
   1. Clustering Benefits – This can group similar users or queries together based on characteristics. By doing this, the company can personalize search results.
   2. Classification Benefits – This will allow the company to assign search questions into predetermined categories. This can help by providing tailored search results as well as help segment search questions into categories.
   3. Association Rule Mining Benefits – This can help identify patterns or relationships within a pool of saved search questions. This can help the internet company provide more targeted results for its user base. Additionally, this technique can provide the internet company the opportunity to understand its user base by collecting data on user experience.
   4. Anomaly Detection benefits – This helps identify unusual behavior that is not seen on a daily basis. This primarily can help with security protocols by seeing where a query is coming from, and from which internet provider the user is connected to.
10. Data Privacy Concerns:
    1. Census data from 1990-1950: I argue this depends on the type of data that is collected in the census survey. If the survey asks for PII, then it is a data privacy concern. If it is just asking high-level questions for a given population, then it is not a data privacy concern.
    2. IP Addresses and visit times of web users who visit your website: No, as this is not PII. However, it is important to know that this can be combined with other information to potentially identify a person.
    3. Images from Earth-Orbiting Satellites: This depends on the location of captured images and its use.
    4. Names and Addresses of people from a telephone book: I would consider this as a data privacy concern because it has PII, but telephone books are seen as publicly available information. If this is a publicly available resource, there are no data privacy concerns.
    5. Names and email addresses collected from the web: This is a data privacy concern.
11. Article Summary: -https://finance.yahoo.com/news/defense-google-flu-trends-142740742.html?guccounter=2

Alexis Madrigals’ article discussed the rise and fall of Google Flu Trends, an experimental tool released by Google in 2008 that aimed to predict the prevalence of the flu based on user searches for flu-related terms. Initially, the tool received positive attention from the media and showed promising results. However, an additional study published in the journal Science criticized Flu Trends for its inaccuracies and highlighted problems with its methodology and search algorithms.

Despite the negative coverage, the article argues that combining Google Flu Trends data with the Centers for Disease Control and Prevention's (CDC) standard monitoring provided better results than either method alone. This highlights the importance of cross-referencing your data or corroborating your findings with another data set.

The creators of Flu Trends had intended it to be a complementary tool, not a replacement for traditional surveillance networks or laboratory-based diagnoses. The article concludes by highlighting the value of combining different data sources and signals to gain a better understanding of flu trends. While Google Flu Trends had its flaws, it still proved useful to researchers and epidemiologists in certain contexts. It is important to understand the purpose of a tool, dataset or methodology if one is trying to extract insights from raw data.