**Q2: Facebook and Google provide many services for free. In return they mine our email and blogs and send us targeted ads. Is that a fair deal?**

Answer:

My personal opinion is it is fair deal. Since, various services are provided by Facebook and Google, from where we get some benefits and easily can use their services when we need. On the other, they are revealing and analyzing our information. It is actually win-win situation. But, both parties should be careful about individual’s privacy. Facebook and Google give lot of terms and condition before using their services. So, we should read those carefully, if you think it could be your privacy issue, then you have choice to deny their service.

2. **Describe the key steps in the data mining process. Why is it important to follow these processes?**

**Answer: CRISP proposed a data mining model which was developed with total six steps.**

* Business Understanding: understanding business domain assist to choose proper data which are required for figuring out hidden insight and knowledge. It is the first phases of data mining.
* Data understanding: various types of data is being generated from various sources daily basis. Data might structure, semi-structured and unstructured data. For data mining, we need to understand data and select those data for mining which will solve our problem and fulfill our purpose.
* Data preparation: since different sorts of data coming from different sources. So, to get accurate result, data should be relevant, clean and high quality.
* Modeling: In this step, model and pattern can be developed from dataset by using different data mining techniques.
* Modeling Evaluation: accuracy, performance and satisfactory level of data mining are measured in this stage.
* Dissemination and rollout: Deployment plan, strategy to monitor and maintain data mining model are developed in this stage. Moreover, final report is created and review of the whole process is done to check any mistake.

This process need to be followed, so that accurate result can be figured out.

**4. Why is data preparation so important and time consuming?**

Answer: for data analysis, the most and significant step is data preparation which involves with cleaning organizing and transform from raw data to certain format that can be easily analyzed. Data preparation is important and time consuming because of large volume data which is generated from different sources. Besides, the type of this generated data is not same. Its type can be different like Structured, semi-structured and unstructured. So, to get accurate insight and result, data should be relevant, clean, consistent and high quality before data analysis.

**6. What are the major mistakes to be avoided when doing data mining?**

1st Mistake: selection of wrong problem for data mining, which results waste of time because of incorrect results.

2nd Mistake: Buried under mountains of data without clear metadata. Instead of digging vast amount of data, it is better to engage with data.

3rd Mistake: conducting data mining without planning which known as actually disorganized Data mining. For instance- over and over again using same algorithm without thinking Next stage, which might be the cause of wasting time and energy.

4th Mistake: Insufficient business knowledge - Lack of knowledge of business domain, which might be reason of meaningless output of data mining.

5th Mistake: Incompatibility of data mining tools and dataset. All tools for data collection, cleaning, mining and representation should work simultaneously.

6th Mistake: paying attention on aggregated output without observing individual records/prediction.

Above mentioned mistakes should be avoided during data mining

**Caselet: Target Corp – Data Mining in Retail**

**2. FaceBook and Google provide many services for free. In return they mine our email and blogs and send us targeted ads. Is that a fair deal?**

In my opinion, I think that is a fair deal, because they provide us many services for free and we’re easy to use it. And the important thing,  they have given the terms and conditions before we use these services if we feel uncomfortable with it, we should not be used their services

**Review questions:**

**2. Describe the key steps in the data mining process. Why is it important to follow these processes?**

* The key steps in the data mining process are:
  + Business Understanding that helps select the right data for pursuing new insights
  + Data Understanding: the data is collected from a lot of unstructured and semistructured sources like images, videos, database… and we need to understand the data for mining
  + Data cleansing and preparation: this step ensures that the data is accurate, complete and consistent. This is an important step
  + Data Modeling: This step is developed using statistical and machine learning techniques to predict result, identify patterns or classify data
  + Model Evaluation: This step is evaluating the accuracy and performance of the data mining.
  + Dissemination and rollout: This step is presented to the key stakeholders and is deployed in the organization

**4. Why is data preparation so important and time consuming?**

* The data preparation is important and time consuming because the data is huge and it should be relevant, clean and high quality
* This is important to assemble a team that has mix of technical and business skills
* The data cleaning can take 60-80% of the time in a data mining
* It includes some functions:
  + Removing duplicate data
  + Filled missing values
  + Data elements should be comparable
  + Continuous values may need to be benned into a few buckets to help with some analyses
  + Outlier data elements need to be removed after careful review to avoid the skewing of results

**6. What are the major mistakes to be avoided when doing data mining?**

There are some major mistakes to be avoided when doing data mining:

* Mistake 1: Choosing incorrect problems for data mining, it will waste of time because we can't get expected results
* Mistake 2: Buried under mountain of data without clear metadata, it's important to interact with the data instead of having a lot of data that we have no method to be able to mine the data
* Mistake 3: Disorganized data mining. If we don’t have clear goals or target, we will spend a lot of time to get meaningless things
* Mistake 4: Insufficient buseness knowledge, if we don't have enough knowledge about business domain, the result will be incorrect and meaningless. So we should have a deep understanding about business domain to understand what we need
* Mistake 5: Incompatibility of data mining tools and datasets, because the output from a tool will be an input of other tools, so all tools should work and be compatible together
* Mistake 6: Looking only at aggregated results and not at individual, because the result can be incorrect at an individual record, so we need to look to both overall and individual
* Mistake 7: Not measuring your results differently from the way your sponsor measures them. If we use data mining for its own sake, we will lose credibility and respect from everyone and more seriously violate the law

**Liberty Stores Case Exercise: Step 3**

**2. What data mining technique would you use to categorize its customers?**

For categorize customers, I think we should use the Cluster Analysis because customers have similar behaviors, preferences, their common wants and needs. Based on this, the Cluster Analysis help identify natural grouping of customers, products, patients and so and it also helps to provide characterization, definition, and labels for populations. Finnally, the result of the Cluster Analysis can be used to segment customers into different categories or groups and used for the targeted marketing or other business applications. Based in the Cluster Analysis, the business can easily understand their customer and develop more effective strategies to provide needs and preferences of their customers

**Caselet 2: FaceBook and Google provide many services for free. In return they mine our email and blogs and send us targeted ads. Is that a fair deal?**  
In my thinking it is like 50-50 deal. Because they disclose our individual privacy level with the services provider. On one hand, Facebook and Google's services can be incredibly useful and convenient, and many people may be willing to trade some of their personal information for the benefits of these services. But finally it's depends on individual values and priorities. And as a user we need to be aware of our data.  
  
**2. Describe the key steps in the data mining process. Why is it important to follow these processes?**

The data mining process typically involves several key steps:

**Business Understanding:** This step to understand the business problem and define the objectives of the domain and determining what data is available, what insights are desired.  
**Data Understanding:** This step is to explore and understand the data like collecting and analyzing data and identify any patterns involves and to be sure the data is related to business domain.  
**Data Preparation:** This step data need to be cleaned formatted and transformed into a format which suitable for data mining, like working on missing data and select relevant data and make data into desired formats.  
**Modeling:** This step is the main step of the data mining process. It's involves selecting appropriate algorithms and techniques to extract patterns and trends from the data.  
**Model Evaluation:** Once patterns are identified, the next step is to evaluate their usefulness and relevance. The evaluation may involve statistical analysis or visualization techniques.  
**Dissemination and rollout:** The final step involves to presenting the solution of the data mining result to the key organization. The insights can be used to make informed decisions, improve processes, and develop predictive models for the business.

The data mining process ensures that the analysis is accurate, reliable, and useful by ensuring that the data is representative, accurate, and relevant, and that the insights are communicated effectively and applied in real-world applications.  
  
**4. Why is data preparation so important and time consuming?**

Data preparation is a crucial step in the data analysis process that involves cleaning, transforming, and organizing raw data into a format that can be analyzed. It is a time-consuming process because the quality of data is directly proportional to the quality of analysis and the accuracy of results. And it's involves some other points like filling missing values, removed duplicate data, data should be comparable, data transformation, data normalization and so on.  
  
**6. What are the major mistakes to be avoided when doing data mining?**

Data mining is a complex process that involves analyzing large datasets to extract useful insights and patterns. Here are some major mistakes to be avoided when doing data mining:

--  Not focusing to identify the problem of data mining. A pointless goal would not good for organization.  
--  Lack of knowledge of business relevant data and metadata. Without focusing lots of data we need to focus to get more business perspective data.  
--  Insufficient business domain knowledge is another major mistake to get insights into the data, it's waste organization valuable time.  
--  Only Looking at aggregated results of data mining and not focusing on individual data.  
--  No having enough knowledge for good combination with data mining tools and dataset.  
--  Disorganized data mining like selecting wrong modeling, without business oriented goal and overall without any proper planning.

2: FaceBook and Google provide many services for free. In return they mine our email and blogs and send us targeted ads. Is that a fair deal?  
  In my opinion, it's our choice. When we create account in FaceBook or Google, there is the term and Privacy Policy that we should read. So we can't say it's fair or not.

2. Describe the key steps in the data mining process. Why is it important to follow these processes?

* Business Understanding: understand the application domain
* Data Understanding: Identify data sources and select target data
* Data Preparation: Pre-process: cleaning, attribute selection
* Modeling: Data mining to extract patterns or models
* Model Evaluation: Post-process: identifying interesting or useful patterns
* Dissemination and rollout: Incorporate patterns in real world tasks

I think data mining is a process to take much time and effort. So these steps are very important to achieve the target of data mining.

4. Why is data preparation so important and time consuming?  
  In my opinion, data preparation is so important because the quality of data is the key to  effectiveness. As we know data mining is success or not based on data preparation. It is time consuming because there are many steps for preparation: Data consolidation, Data Cleaning, Data Transformation, Data Reduction. With each steps, there are 3 sub steps. In order to get well-formed data, data should be spent into many main steps and sub-steps.

6. What are the major mistakes to be avoided when doing data mining?

* Selecting the wrong problem for data mining
* Ignoring what your sponsor thinks data mining is and what it really can/cannot do
* Not leaving insufficient time for data acquisition, selection and preparation
* Looking only at aggregated results and not at individual records/predictions
* Being sloppy and keeping track of the data mining procedure and results
* Ignoring suspicious (good or band) findings and quickly moving on
* Running mining algorithms repeatedly and blindly, without thinking about the next stage
* Naively believing everything you are told about the data
* Naively believing everything you are told about your own data mining analysis
* Measuring your results differently from the way your sponsor measures them

Liberty Stores Case Exercise:  
2. What data mining technique would you use to categorize its customers  
I think Artificial Neural Networks (ANN) is the technique using to categorize the customers. The neural network can be trained by making a decision over and over again with many data points. It will continue to learn by adjusting its internal computation and communication parameters based on feedback received on its previous decisions.