BI Assignment

ISDS 4180 – Fall 2016

1. **What is the difference in efficiency and effectiveness? Be sure to supply an example.**

   The fundamental difference between efficiency and effectiveness lies in the intended results of the two concepts. Effectiveness is present when the activities being performed are correct and will ensure that the underlying goal of the endeavor is reached. Efficiency differs because it is only present when the activities are being performed well and are structured in a way to ensure the relative input needed to achieve the desired output is optimized. Ideally, organizations would like to achieve maximum effectiveness and efficiency possible in all endeavors, but this is much simpler in theory than in practice. For example, let us assume that we are working in the IT security team at a multi-billion dollar corporation with around 5,000 employees, and our goal is to optimize our efficiency and effectiveness. However, we are constrained on cost and headcount. If our security stack has 12 different security products, we might choose to divide up the responsibility for managing each of these twelve products into three groups by skills needed and the tools’ functions. These three groups could be threat management software, network security tools, and endpoint management solutions. Then, we would give one group to each of our three IT security team members. This would minimize the learning curve since the skills needed for the tools in each group are related. This exercise would also allow each team member to be efficient in their daily work as well as project work since the tools they are working with are all related to one another. However, this would not necessarily ensure that their work is effective. The overarching goal to an effective IT security team is to keep the enterprise and all of its data secure. To ensure they are effective, the team members would have to work together to understand the connections between each of their three areas, ensuring no malicious operations are “slipping through the cracks”.

2. **How has the advancement in computers, from the mainframe in the 1950’s to the microcomputers and networks in the 1990’s, helped evolve Business Intelligence?**

   Initially, technology began to take root in business with the mainframe computers of the 1950’s where the primary use of the computational technology was to reduce the cost of transactions by automating them at an operational level. Then, in the 1970’s and 1980’s, technological support became possible for other areas of the business such as human resources and marketing. Productivity was improved with systems that aided manufacturing and material requirements planning, and efficiencies were found by using point of sales technologies. These tools were primarily helpful to operational and functional management. However, the transaction processing technologies did not enhance the work of middle-level management. It was during this time that management information systems were introduced as an additional layer to companies’ technology stacks, enabling reporting and analysis capabilities. This shift in thinking was still aimed to improve efficiency, but it seems to be one of the first forays in using technology to drive more intelligent decision making. The microcomputers of the 1980’s enabled managers with a channel through which they could find answers in regards to the effectiveness of different business areas. The late 20th century also saw the first defined concepts around a major decision support system and connections between computational methods and electronic data processing. Furthermore, the 1990’s was when we saw a greater prevalence of networks, leading to group support and integrated computer systems that would go on to enable organizations to share knowledge and better manage customers. Where previously technology could only be leveraged to ask questions about efficiency, these decision support systems helped business make great strides toward answering those effectiveness
Taylor C. Veillon

questions. These advancements started a chain reaction that has enabled the business world to build a platform of technology to stand upon. This platform provides organizations with the information they need to understand what strategic changes need to take place and with the capabilities to make those changes a reality. Though IT and functional areas still often operate separately, organizations are leveraging technology to help make more intelligent decisions about business.

3. What two technological advances, particularly occurring during the 1990s, have contributed to closing the gap between “Business” and “Intelligence”? Be sure to explain how.

The first technological advance of the 1990s that helped close the gap between business and intelligence was the increased prevalence of the internet. It provided a common platform through which information could be shared among different areas of an organization. The second technological advance involved disk storage technologies, or more specifically redundant array of inexpensive/identical disk (RAID) technology. This innovation allowed enterprises to store historical data efficiently, enabling these organizations to maintain their own source of long-term data to understand patterns and trends. The widespread use of the growing internet helped to close the gap between business and intelligence, especially in regards to technology, by changing strategic management’s focus within their perspective organizations, helping them to identify new strategic opportunities, and fundamentally altering the business of business. The internet helped to close the gap by making it easier for functional and technical areas to interact, providing more business knowledge to the data being analyzed. The increasingly less expensive cost of mass storage devices set off a chain reaction of innovations that continue to alter the way we use technology to be more intelligent about how we do business even to this day. Initially, the easier storage options allowed for quicker and more effective backing up of enterprise data. They also introduced the capability for data warehousing, removing taxing analysis processes from the protected transactional systems, and eventually led us to development of much more robust and user friendly data mining, data analysis, and data visualization tools. The advanced tools we work with today enable decision makers to intelligently interact with the data and models used. These techniques pooled together have helped transform data into business intelligence. Easier storage options eventually led to the development of tools that made the intelligence easier to find, making it possible for more business stakeholders to be involved in the technical work, further closing the gap.

4. What are the key ingredients necessary for making effective decision? Explain why each ingredient is necessary.

The first key ingredient is to define a set of goals that the entity will work to achieve. The second key ingredient is that there must be a metric defined to measure whether or not a set of activities is shifting the organization toward or away from the set of defined goals. The final ingredient is that information from intermediate results and feedback must be provided to decision makers in time for them to change direction if needed. The first step of defining specific goals helps to set the organization up to succeed, ensuring that all stakeholders are on the same page regarding the organization’s direction. Effective goals contain defined success targets and have the means through which progress can be measured. Once an organization understands its goals, it must start making decisions and moving toward achieving them. Whether or not certain activities or decisions move us toward or away from a goal is determined by our method for measuring progress. This component is vital to ensure stakeholders have the information they need to make a decision to adjust certain activities that may be pulling us away from a goal, potentially reducing our effectiveness. Finally, feedback about progress must be provided to the right people in a timely manner. If an activity is performed or a decision is made that may be to the detriment of the organization, course corrections can be made easily if the communication
channel is already established for this type of feedback. Delays in relaying such information can be costly to the organization in both the short and long-term.

5. **Write an effective decision making goal.**

   “Taylor Veillon will complete all seven courses included in the “Ethical Hacking Bundle” by April 30, 2017. Taylor is to complete one course per month over the next seven months and provide evidence of completion as well as performance results from each course to her manager by the last day of each month.”

   This goal’s specific target or measure of success is the completion of all seven courses included in the ethical hacking bundle by the deadline provided. The means to measure progress are the reports of whether or not the required course is completed each month as well as a report of my performance to provide further detail about how well I am mastering the material. I wrote this goal with the S.M.A.R.T. framework in mind, ensuring it was specific, measurable, attainable, realistic, and timely.

6. **What is Business Intelligence?**

   Business intelligence (BI) is the method through which we can obtain and distribute the information needed to maintain that important feedback loop that is our third key ingredient for effective decision making. BI is the discipline that enables us to produce accurate and valuable information and deliver it timely to the correct stakeholders to aid them in making more effective decisions. Whether it is forecasted values, spreadsheets of financial data, or reports with visualizations, one of the most important characteristics about business intelligence is that it can be understood by the decision maker trying to use it. The “intelligence” has to be presented in a way that is meaningful to the target stakeholder and draws clear connections between the data showcased and the information that can be gleaned from it. In addition to technical understanding of database structure, data mining, and modeling techniques, an analyst looking to present BI deliverables needs to have an understanding of the data. It is that understanding that takes the data from raw numbers to high-value pieces of intelligence and insight into the business. This is the information the stakeholders can use to make decisions.

7. **What is the difference between “layout-led discovery” and “data-led discovery” queries? Find an example of each on the web and include an image for each one (ensure the image is embedded into the Word document).**

   The underlying differences between layout-led discovery and data-led discovery are whether or not we know what intelligence we are looking for, where to find it, and how the intelligence is unearthed within the business intelligence deliverable itself. Layout-led discovery is the most common form of business intelligence where analysts use reports to deliver the requested “intel”. In this form of discovery, we know what question we are trying to address, and we have a good prediction of what sources we need to use to compile the data. However, with layout-led discovery, the stakeholder is limited to only the information the analyst thought to include. There is no way to drill into greater detail unless those additional data points were included when the report was constructed. We are constrained by the layout. Data-led discovery is much more interactive than layout-led discovery. Additionally, data-led discovery differs from layout-led discovery in that the analyst may know the question being asked but may not know what data sources to search to find the answer. Typically, as more initial information is found in regards to the original question asked, the gained insight alters what the business thought they were looking to find. Even if the analyst knows what data sources to use, the data sets may be too large or provide too much irrelevant detail. In this case, the analyst would want to start at a higher level of detail and allow the business user to drill down into more specific data points as they see fit. This is where the interactivity of the BI deliverable comes into play. It
is important that the analyst create an environment that the business users can easily manipulate to find the intelligence they are looking for.

The image below actually showcases both layout-led discovery and data-led discovery. On the conventional branch, we see our layout-led discovery example: a traditional SQL query being used to obtain an output of records. On the bottom branch, we see an example of a more interactive, data-led discovery deliverable where users can slice and dice the data as they see fit on an interactive web interface.

The image below contains a very advanced example of data-led discovery. Exabeam is a security software vendor, and this tool enables security team members to actively search logs and enterprise endpoints to “hunt” for various types of threats, allowing them to filter by their defined requirements. Additionally, it updates its interface frequently to show various threats detected and flagged. Within each of the events reported, users can drill down into farther detail, even going so far as to view the actual logs the data is being pulled from.
8. **What is latency and explain how latency in the data is important/not important with upper management (strategic planning), middle management (tactical planning), and lower management (operational planning)?**

Latency, in regards to business intelligence, measures the delay between the exact time a transaction is recorded in an organization’s transaction processing system and when that data point is loaded into the business intelligence or analysis system. Upper management stakeholders who are focused on the organization’s strategy and ensuring it is headed in the right direction to achieve its goals do not need as much detail in their business intelligence deliverables as other levels of management. They use business intelligence to spot or predict long-term trends that may indicate if strategic adjustments need to be made. Because they need less detail and are looking for longer-term information, they can have more latency. In other words, the time between daily operational occurrences and when upper management needs to see those data values reflected in business intelligence deliverables is longer. Latency is not important when it comes to upper management as long as they are able to see important long-term trends in a timely manner. These stakeholders would probably be satisfied with layout-led discovery deliverables.

For middle managers, greater detail is needed than upper management, but they still are not “within the weeds” of every day occurrences. Their goals are short-term, but they are still doing the planning for their relative areas. Therefore, they need summarized business intelligence that still allows them to drill into more detailed information if needed. These stakeholders are ideal candidates for interactive, data-led discovery deliverables. Because they need more up-to-date information than upper management but still are not concerned with day-to-day transactions, they need less latency, or less delay between the time the transaction occurs and when they see that data point. However, some latency is still acceptable, so with middle management, the latency of the data they are looking for is minimally important.

Finally, for lower management, their goals are focused on daily operational milestones and planning in the short-term. These stakeholders need minimal latency and a highly
responsive and interactive business intelligence system. They are also ideal candidates for a
data-led discovery environment, so that they can react quickly to changes and drill down into as
much detail as necessary to make adjustments on a daily or even hourly basis. It is important
that there is as little latency as possible in the data lower management needs to use.

9. Discuss at least four differences between Online Transaction processing systems (OLTP) and
Online Analytical Processing systems? Why are OLAP systems better than OLTP systems for
Business Intelligence?

Four differences between online transaction processing systems (OLTP) and online
analytical processing systems (OLAP) are in regards to the purpose, structure, frequency of
updates, and data storage of the two different types of systems. An OLTP’s purpose is to process
and store an organization’s day-to-day transactions efficiently. In contrast, an OLAP system’s
purpose is to be a source for business intelligence, not to keep up with the everyday operations
and transactions of an organization. Another key difference between OLTP and OLAP is the
structure of these systems. OLTP systems are stored in third normal form which keeps the data
separated in a relational database structure, prevents data redundancy, and makes records
easier to update. This structure enables these systems to process many transactions at a time.
OLAP systems have a different structure. The relational database structure is often still
maintained, but the structure is a bit more flattened to require less table joins for analyzing the
data. Because they are designed for speed of access to data, they are often denormalized, which
means that data redundancy is acceptable. The normalized design of OLTP systems is replaced
with star and snowflake schemas where the data is organized by separating fact and dimension
tables. OLTP systems are updated as often possible and store transactions as they occur, so the
data is real-time or near real-time. In contrast, OLAP systems do not reflect the data in real-
time. Data is pulled from OLTP systems, and records are updated in OLAP systems at set
intervals determined based on latency tolerance. The fourth difference is in regards to how
much historical data is stored. OLTP systems use an archive process to maintain efficient
processing. Since these systems are more concerned with day-to-day operations, they do not
need to hold historical data and may sometimes only represent the current value of a data
point, storing no past values. OLAP systems, however, look to keep large sets of historical data.
This historical data allows users to look for long-term trends and patterns.

In addition to having more historical data and the taxing process of running large
analyses not being an issue, OLAP systems are better for BI than OLTP systems primarily because
of the structure and quality of the data. Through the extract, transform, and load process, data
is pulled from OLTP systems and loaded into OLAP systems. During the process, the data is
consolidated, cleansed, and then pieced together into a star or snowflake schema from which it
is more ideal to run analyses.

10. Define data mining and give an example of how a company may use data mining techniques?

When looking to discover knowledge within the enterprise data of an organization, data
mining is considered the analysis step within the discovery process. It is the practice of using
statistical methods with algorithms to dig through datasets to unearth interesting patterns or
tendencies of the data. “Data miners” start at the lowest level of detail or most granularity and
work to build models using applied statistical methods to discover any hidden commonalities
within the data. For example, an oil and gas company that has well pads producing crude oil
could use data mining techniques to build models to help engineers monitor the massive inflow
of production data. The models could be built to help predict if a certain piece of equipment (i.e.
valve, tubing, or crank on a pump jack) may fail or leak before its typical maintenance and
replacement cycle. Using predictive techniques like linear regression, an analyst could build a
model using certain defined predictors like production volume, pressure, and days since
maintenance to predict if key pieces of equipment might need to be more closely monitored or replaced. This could help by reducing the costs of having to react to a failure and deal with shut-ins and longer downtime and create a more intelligent preventative maintenance cycle.

11. **Discuss the difficulties of using transactional data for business intelligence?**

One of the difficulties that organizations encounter when trying to use transactional data for business intelligence is that it is typically unlikely for all of an organization’s transactional data to be in one place. Many organizations have multiple transaction processing systems for needs such as manufacturing, accounting, and order processing, and determining what data sources are needed for analysis purposes can be very difficult. Additionally, the measures that are often found for BI insights are brought together using data points from systems all across an enterprise, so anytime a user needs to calculate a measure, all of this data would have to be combined. Finally, a last difficulty is also related to the need for data from multiple systems to be combined. Different systems typically have different numbering schemes and codes for the same items, so connecting different data sources will require the analyst to find commonalities between different schemes.

12. **Define a measure, a dimension, an attribute, and a hierarchy? Give an example of each.**

I would like to speak about a database for the LSU Olinde Career Center that has an interview fact table that includes a student ID, company ID, recruiter ID, and date ID that connects the main fact table to four dimension tables for students, companies, recruiters, and dates.

A measure details an aspect of an organization’s performance through a numeric quantity. Information gleaned from measures is often used to evaluate performance and support a decision being made. Measures are the facts used for the foundation of the business intelligence an analyst provides. A measure in the case of my example would include a count of how many interviews the career center has conducted. A dimension helps us categorize a single point of data that we have with an aggregate measure. An example of a dimension would be choosing to see our count of interviews by year, perhaps to see if the volume went up between 2015 and 2016 after the career center started a new student outreach program. Attributes are extra pieces of information that relate to a dimension member. These values cannot be the unique identifier or the description of a dimension member. Examples of attributes for the student dimension would be major, classification, and gender. Finally, a hierarchy is made up of two or more levels within a dimension. An example of this could be found within the company dimension table. The hierarchy structure could be industry sector, industry, and then company name.

13. **Discuss the difference between a star vs a snowflake schema?**

The difference between a star and snowflake schema is how hierarchies are represented within the design of the database. Hierarchies provide different levels of detail within the measures of a data mart. In a star schema, all levels of a hierarchy is stored in the same dimension table. In contrast, within a snowflake schema, each level of a hierarchy is placed in its own dimension table.

14. **What does ETL stand form? State what happens during each step of process (what happens during the ‘E’, during the ‘T’, during the ‘L’)? Be sure to give an example for each.**

ETL stands for Extract, Transform, and Load, and this is the process through which data is pulled from transaction processing systems and inputted into a data mart. The process also includes a data cleansing step to ensure that the data is of high quality. The first step of the process, extract, involves copying the data from one or many transaction processing systems.
This phase’s goal is to convert the data extracted into a consistent format that can then be transformed. I would like to use the Chimes restaurant here in Baton Rouge as my example of an organization using the ETL process. For Chimes, the extract phase might occur each night after the restaurant closes and would involve pulling the data out of the transaction processing system that records each customer purchase as well as perhaps an order management system that tracks the timing of each order and subsequent delivery. The extract process would need to ensure these two sources of data are prepared for transformation. The next step is the transform phase of the ETL process. During transformation, a series of rules might be applied to the extracted data to ensure it passes quality checks and is properly joined. For instance, each order should correspond to a purchase or multiple purchases, and the transform phase would ensure that the relationship between both sets of data is defined. Finally, the ETL process finishes by loading the data into the end target system. In the case of Chimes, this may be a data mart they use to run reports for inventory management, to determine what dishes are favored, and to understand what days are “slow”. If the load process puts the data in a database, it must handle the constraints of the destination database’s design (i.e. required fields and uniqueness). The data could also just be outputted into a flat file to be loaded into Excel and analyzed on a smaller scale.

15. What is a data mart?
A data mart is an intermediate container to store data outside of OLTP systems. A data mart holds historical data and is not involved in the day-to-day operations of an organization. Data marts are specifically designed to be used to create business intelligence. They are structured so that data can be accessed quickly, storing the data in either a star or snowflake schema for less table joins when running reports and performing analyses. Data marts are made up of measures, dimensions, attributes, and hierarchies.

16. If we can incorporate data directly from OLTP systems, discuss why do we need data marts?
There are three key reasons that comes to mind when I think about the importance of data marts even though we can technically incorporate data directly from OLTP systems. The first reason is to not strain the organization’s key transaction processing systems. Analyses can be very taxing on processing power, and it is important to interfere with this system as little as possible. The second reason is due to the ETL process that loads the data into a data mart from an OLTP system. During this process the data is cleaned, reducing the likelihood of inconsistent values or errors that can be typically found in transactional data. This helps ensure better data quality for our analyses. The last reason is due to the structure of this data container. Not only can we combine multiple transactional data sources into one data mart, the design of the storage itself enables us to calculate complex measures and slice and dice information based on a variety of attributes.

17. What three questions should decision makers answer to help design a data mart? Explain why these three questions are important and how they are used to shape the data mart.
The three questions decisions makers need to answer to help design a data mart include the following. What facts, figures, statistics, and so forth do you need for effective decision making? How should this information be sliced and diced for analysis? What additional information can aid in finding exactly what is needed? These questions are important to ask of our decision makers because involving them in the design process creates two key results. First, stakeholders who are allowed to contribute feel ownership in a deliverable, and since they do determine whether or not a BI project is successful, it is imperative that the analysis team has their support. Second, since the key decision-makers are the ones “in the weeds” of their specific business area and will be the ones using the BI deliverable, it is to the analyst’s
advantage to really involve them in the process of choosing the right data and how it should be presented. The answers to these questions are also very impactful on the overall data mart’s creation. Our first question about needed facts will be compared to what information is available in the OLTP data to define what measures we need to include in the data mart. Our measures make up our fact tables and are fundamental to the structure of our data mart and what business intelligence we can glean from it. The second question about how we want to slice and dice the data can help us understand what dimensions we want to add to our data mart and how those dimensions are connected to our fact tables. Finally, the third question can help us define any attributes that might be necessary to provide additional information about a dimension and further assist users in finding the intelligence for which they are looking.