UNIVERSIDAD CATOLICA  
“NUESTRA SEÑORA DE LA ASUNCION”  
CAMPUS ALTO PARANA  
FACULTAD DE CIENCIAS Y TECNOLOGIA

Iván Ariel Cáceres Cañete  
Ariel Hernán Landaida Duarte

**DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY**

UNIVERSIDAD CATOLICA  
“NUESTRA SEÑORA DE LA ASUNCION”  
CAMPUS ALTO PARANA  
FACULTAD DE CIENCIAS Y TECNOLOGIA

Iván Ariel Cáceres Cañete  
Ariel Hernán Landaida Duarte

**DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY**

Ricardo Luis Brunelli Montero, Ing.  
Tutor

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

Proyecto de Fin de Carrera presentado como requisito parcial para optar al título de Licenciado en Análisis de Sistemas.

Facultad de Ciencias y Tecnología, Universidad Católica “Nuestra Señora de la Asunción”

Tutor: Ing. Ricardo Luis Brunelli Montero

Landaida Duarte, Ariel Hernán; Cáceres Cañete, Iván Ariel. (2016); DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY, una aplicación para ayudar a las personas a analizar, visualizar y compartir información rápidamente. Hernandarias, Universidad Católica. 110 p.

Tutor: Ing. Ricardo Luis Brunelli Montero.

Defensa de Proyecto de Fin de Carrera.

Palabras clave: Data Discovery, Business Intelligence.

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

Proyecto de Fin de Carrera presentado como requisito parcial para optar al título de Licenciado en Análisis de Sistemas.

Mesa Examinadora

Prof. Nélida Elizabeth Delgado, Lic.  
Presidente de Mesa

Prof. Manuel Chamorro Alderete, Ing.  
Miembro de Mesa

Prof. Ricardo Luis Brunelli, Ing.  
Presidente de Mesa

Nota obtenida:

DEDICATORIA

A mis padres, por la oportunidad de existir, por su sacrificio en algún tiempo incomprendido, por su ejemplo de superación incansable, por su comprensión y confianza, por su amor y amistad incondicional, por los consejos que siempre me han dado

para sobrellevar los desafíos en la vida.

A Dios por la fortaleza que siempre nos ha dado en todos los momentos de

nuestra vida.

Gracias, a nuestro tutor, el Ing. Ricardo Luis Brunelli. Gracias por su paciencia,  
dedicación, motivación, criterio y aliento. Ha hecho fácil lo difícil. Ha sido un  
privilegio poder contar con su guía y ayuda.

A nuestra familia quienes por ellos somos lo que somos. A nuestros padres por su  
apoyo, consejos, comprensión, amor, ayuda en los momentos difíciles, y por  
ayudarnos con los recursos necesarios para estudiar. Nos han dado todo lo que  
somos como persona, valores, principios, carácter, empeño, perseverancia, y  
coraje para conseguir nuestros objetivos.  
  
Ivan Ariel Caceres Cañete

A Dios por la fortaleza que siempre nos ha dado en todos los momentos de

nuestra vida.

Gracias, a nuestro tutor, el Ing. Ricardo Luis Brunelli. Gracias por su paciencia,  
dedicación, motivación, criterio y aliento. Ha hecho fácil lo difícil. Ha sido un  
privilegio poder contar con su guía y ayuda.

A nuestra familia quienes por ellos somos lo que somos. A nuestros padres por su  
apoyo, consejos, comprensión, amor, ayuda en los momentos difíciles, y por  
ayudarnos con los recursos necesarios para estudiar. Nos han dado todo lo que  
somos como persona, valores, principios, carácter, empeño, perseverancia, y  
coraje para conseguir nuestros objetivos.  
  
Ariel Hernán Landaida Duarte

RESUMEN

Hasta hace poco tiempo, la mayoría de las organizaciones proveían datos es­tructurados, limpios, e integrados, resumidos a niveles convenientes para plata­formas convencionales. Datawarehouse y BI (Business intelligence) dominaban ese enfoque. Otras organizaciones, principalmente aquellas centradas en internet, desarrollaron algunas alternativas para gestionar y analizar grandes volúmenes de datos directamente de sus sitios y aplicaciones web, hoy generalmente denomina­do Big Data. Aquellos datos obtenidos, en su mayoría, eran heterogéneos y hasta inclusive no estructurados, y esa situación generó la necesidad de crear otro tipo de herramienta que ayude al tomador de decisión en la búsqueda de patrones y relaciones. Este nuevo enfoque, denominado Data Discovery, no podía ser igual a las técnicas ya tradicionales, además debía tener características como innovación visual, facilidad de uso, UX (User Experience) para que se asemeje a un BI guiado por un usuario experto del negocio. En este trabajo se presenta una propuesta del estado del arte del área de BI y específicamente Data Discovery. Se aplican estas técnicas a datos de dos instituciones del estado, demostrando los beneficios de aplicar este tipo de técnica.

Palabras clave: Data Discovery, Business Intelligence, Datawarehouse.

ABSTRACT

Until recent time, most organizations provide structured, clean, and integrated data, summarized in desirable levels for conventional platforms. Datawarehouse and BI (Business intelligence) ruled that approach. Other organizations, mainly those focusing on the Internet, developed some alternatives to manage and analyze large volumes of data directly from their websites and web applications, now gene­rally called Big Data. Those data, mostly, were heterogeneous and even including unstructured, and this situation generated the need to create another type of tool that helps the decision maker in the search for patterns and relationships. This new approach, called Data Discovery, could not be equal to the traditional techniques also should have features like visual innovation, usability, UX (User Experience) so that it resembles a BI guided by an expert business user. This paper presents a proposal of state of the art area BI and Data Discovery is presented specifically. These techniques are applied to data of two state institutions, demonstrating the benefits of applying this kind of technique.

Keywords: Data Discovery, Business Intelligence, Datawarehouse.

### **INDICE**

1. [CAPITULO 1 1](#bookmark4)
   1. [Introducción 1](#bookmark5)
   2. [Planteamiento del problema 1](#bookmark6)
   3. [Objetivos 2](#bookmark7)
      1. [Objetivo General 2](#bookmark8)
      2. [Objetivos Específicos 2](#bookmark9)
   4. [Justification 3](#bookmark10)
2. CAPITULO 2 3
   1. MARCO TEORICO 4
      1. [Business Intelligence 4](#bookmark11)
      2. [Data Discovery Analysis 13](#bookmark16)
      3. Framework de Análisis de Negocio de Gartner 15
3. [CAPITULO 3 24](#bookmark23)
   1. [Selection de la herramienta para Data Discovery 24](#bookmark24)
      1. [Cuadrante Mágico de Gartner 24](#bookmark25)
   2. [Aplicacion de Data Discovery a datos de instituciones del Estado . 29](#bookmark26)
      1. Datos de la ANDE y de la DGEEC 29
      2. [Dashboard de control / monitoramiento 30](#bookmark27)
4. CAPITULO 4 46
   1. [Marco Metodologico 46](#bookmark37)
      1. [Alcance 46](#bookmark38)
      2. [Enfoque 46](#bookmark39)
      3. [Tecnica e Instrumentos de recoleccion de datos 46](#bookmark40)
5. CAPITULO 5 46
   1. [Conclusiones y Trabajos futuros 47](#bookmark41)

Referencias . . . . . . . . . . . . . . . . . . . . . . . 50

### **LISTA DE TABLAS**

2 1. Técnicas actuales de BI

### **INDICE DE FIGURAS**

1. Etapas de BI como fuentes de datos de calidad .... 5
2. [Framework Analitico de Negocios Gartner 16](#bookmark17)
3. Espectro Analitico 17
4. Tipico uso de estilos analiticos 19
5. Características de BI en niveles y Plataforma Analítica . 23
6. Cuadrante MAgico para BI y Plataformas Analíticas . . 25
7. Puntuaciones de producto o servicio para analisis descen­tralizado 26
8. Puntuaciones de producto o servicio guiados por Data Dis­covery . . . . . . . . . . . . . . . . . . . . 27
9. Arrastre el campo país para el campo desplegable señalado. 28
10. Arrastrar hojas de trabajo al dashboard 29
11. [Clientes Facturados vs Crecimiento Poblacional .... 31](#bookmark28)
12. Clientes Facturados vs Crecimiento Poblacional .... 32
13. Clientes Facturados vs Crecimiento Poblacional .... 32
14. Clientes Facturados vs Crecimiento Poblacional .... 33
15. Clientes Facturados vs Crecimiento Poblacional .... 34
16. Clientes Facturados vs Crecimiento Poblacional .... 34
17. Proyección de Clientes Facturados Crecimiento Poblacional

Y Consumo De Energía para los Próximos 5 Anos ... 35

1. [Energía Consumida vs Importe Facturado 36](#bookmark29)
2. Energía Consumida vs Importe Facturado 37
3. Estadística de consumo de electricidad por sector(1990-2014) 38
4. Importe facturado por ano y sector(1990-2014) .... 39
5. Tasa de crecimiento vs Consumo de energía 40
6. Tasa de crecimiento vs Consumo de energía, filtrado por el

departamento Alto Paranía . . . . . . . . . . . . 41

1. Proyecciín de clientes y consumos para los próximos 5 anos 42
2. Panel de Tasa de crecimiento poblacional y Consumo de

energía anual . . . . . . . . . . . . . . . . . 42

1. [Consumo por departamento 43](#bookmark35)
2. Proyección de crecimiento poblacional y consumo de energía...…44

CAPITULO 1

1. Introducción

En este trabajo fue elaborada una propuesta de estado del arte del área de BI (Business intelligence) y específicamente Data Discovery. El objetivo fue apli­car los conocimientos adquiridos al entorno nacional, en este caso aplicados a una institución del estado (Administración Nacional de Electricidad, ANDE), para el apoyo en la toma de decisiones. Aplicaremos las técnicas de Data Discovery a los datos de dos instituciones del estado, específicamente la Administración Nacional de Electricidad (ANDE) y Dirección General de Estadísticas, Encuestas y Cen­so (DGEEC), donde demostraremos que con datos de calidad podríamos detectar oportunidades que nos faciliten la toma de decisiones en la institución. Utiliza­remos conjuntos de datos de las instituciones mencionadas mías arriba para este fin. Este trabajo está estructurado de la siguiente forma: El capítulo I presenta el planteamiento del problema, los objetivos generales y específicos, concluyendo con la justificación. El capítulo II presenta el marco teórico. Se compone de con­ceptos, componentes, tecnologías de BI, y el framework de Gartner. El capítulo III presenta los diferentes paneles indicadores (dashboards) realizados con la he­rramienta seleccionada para aplicar data discovery sobre los conjuntos de datos existentes. Se explica el cuadrante mágico de Gartner, el cual apoyo a la selección de la herramienta mencionada. Además, se presentan ejemplos de análisis de da­tos y la información obtenida, como resultado de la combinación del uso de las herramientas y datos.

1. Planteamiento del problema

Actualmente la mayoría de las instituciones públicas e incluso las del sector pri­vado, tienen un bajo nivel de inversión en tecnología. Generalmente en empresas que exigen toma de decisión, con frecuencia optan por decisiones de negocios no óptimas debido a que no poseen la suficiente experiencia o suficientes datos procesa­dos del negocio para llegar al correcto análisis, o pueden estar usando herramientas incorrectas. Teniendo únicamente la experiencia como herramienta, puede ser su­ficiente sólo en organizaciones pequeñas, donde se adquiere conocimiento sin necesidad de alguna herramienta de análisis de datos, por ejemplo, sabríamos cuales son los productos más vendidos o más rentables en un negocio pequeño, y esto se complica a medida que la organización crece, cuando la cantidad de sucursales y variedad de productos es amplia, se pierde el control sin ayuda de estas herramientas.

La calidad del conocimiento se basa principalmente en la calidad de la información. Estas informaciones son obtenidas a través de un análisis profundo de datos, por consiguiente, estos datos también deben ser de buena calidad. Si una organización posee la capacidad de obtener lo que desea por medio de los datos, es seguro que su crecimiento será positivo, debido a que tomará mejores decisiones, y estos ayudarían a su evolución y estabilidad a lo largo del tiempo.

Existen ocasiones en que la organización posee suficientes datos, pero no con­sigue analizarlos o procesarlos por falta de conocimiento del negocio, además de desconocimiento técnico de las herramientas.

1. Objetivos
2. Objetivo General

En este trabajo se aplican las técnicas de Data Discovery a datos de dos ins­tituciones del estado: ANDE y DGEEC. Se proponen algunos tableros de control (Dashboards) con indicadores, para demostrar que con los datos proveídos es po­sible detectar situaciones, y estimar o visualizar acontecimientos de interés para una organización.

1. Objetivos Específicos

* Estudio del estado del arte de BI y Data discovery.
* Selección de la herramienta adecuada para este trabajo.
* Análisis y cruzamiento de datos para el descubrimiento de situaciones de interés.
* Elaboración de dashboards que reflejen el análisis de los datos con la herra­mienta seleccionada.
* Elaboración de gráficos de tendencias, proyecciones de consumo y clientes de la institución.

1. Justificación

Falta de un ambiente analítico corporativo, que proporcione informaciones e indicadores necesarios para la toma de decisión. Falta de agilidad para elaborar informes con indicadores de tendencias de consumo de energía, que apoye a la planificación territorial de expansión de la red de transmisión eléctrica. Falta de indicadores de consumo geográfico de electricidad, relacionados con indicadores poblacionales, para apoyo en la planificación de inversión en el aumento de capacidad de transformadores por región.

CAPITULO 2

1. MARCO TEORICO
2. Business Intelligence

Business Intelligence (BI) o Inteligencia de negocio, es definido como la capa­cidad de una organización para tomar todos sus procesos y capacidades, y luego convertirlos en conocimiento, o, en otras palabras, obtener información correcta pa­ra la persona correcta, en el tiempo correcto, a través del canal correcto (Kumari, 2013).

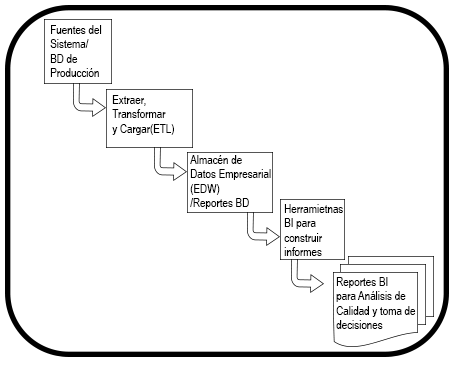
Esto produce grandes cantidades de información que pueden llevar al desarrollo de nuevas oportunidades para la organización. La identificación e implementación de estas oportunidades en una estrategia efectiva, puede proporcionar una ventaja competitiva de mercado y la deseada estabilidad a largo plazo (Rud, 2009).

Las tecnologías de BI ofrecen vistas históricas, actuales y predictivas de las ope­raciones de una empresa. Algunas de las funcionalidades más comunes de BI son reportes, procesamiento analítico en línea (en adelante OLAP, Analytical Proces­sing), minería de datos, minería de procesos, procesamiento de eventos complejos, gestión del rendimiento empresarial, benchmarking, minería de texto, análisis predictivos y análisis prescriptivos. BI tiene como objetivo apoyar una mejor toma de decisiones, por tanto, un sistema de BI puede también ser llamado sistema de apoyo a las decisiones o DSS (del inglés, decisión support system) (Rud, 2009).

Las herramientas de BI son diseñadas para obtener, analizar y reportar datos. Estas herramientas generalmente leen datos que han sido previamente almace­nados, frecuentemente, aunque no necesariamente en un almacén de datos (en adelante, datawarehouse). BI ha experimentado un alto crecimiento y ha ganado mucha popularidad. De acuerdo a (Hancock y Toren, 2006), BI es “un conjunto de conceptos, métodos, y tecnologías para convertir datos separados o aislados de una organización en información útil para mejorar el rendimiento del negocio “

En un ambiente de BI, se extraen datos de diferentes fuentes, se transforman y posteriormente se cargan (ETL, extraction - transformation - load) dentro de un datawarehouse y desde este repositorio son utilizados para generar reportes trans­versales a toda la organización. El proceso de BI y sus varias etapas son mostrados en la Figura 2.1. La calidad de los datos juegan un rol crítico e importante en el éxito de la inteligencia de negocios, ya que la mala calidad de los datos pueden afectar las decisiones de negocio en todos los niveles de la organización, además de impedir el crecimiento de la organización (Kumari, 2013).

Figura 2.1: Etapas de BI como fuentes de datos de calidad. Kumari, N. (2013)



Una cuestión fundamental es el hecho que una organización corre sobre datos; y actúa como insumo para el motor de la industria corporativa. Una organización no puede comprender a sus clientes, proveedores, competidores o a su propia gen­te, procesos, y rendimiento sin datos de buena calidad. Por consiguiente, la alta dirección de una empresa y el área de TI (tecnología de la información) deberían trabajar juntos para asegurar datos de alta calidad (Eckerson, 2009).

1. Componentes de BI
2. OLAP: OLAP se refiere al mecanismo por el cual los usuarios de una organización pueden explorar y realizar cortes de datos, usando herramientas sofisticadas que permiten la navegación de dimensiones tales como el tiempo o jerarquías. OLAP, provee vistas resumidas multidimensionales de datos del negocio de una organización, y es usado para reportes, análisis, modelado y planificación para la optimización de una organización. (Malhotra, 2001)

Las técnicas y herramientas OLAP pueden ser usados para trabajar con datawarehouse o con datamarts (un subconjunto de datos de un área específica) diseñados para sistemas sofisticados. Este tipo de consultas son requeridas para descubrir tendencias y analizar factores críticos. Los reportes generan vistas agre­gadas de datos para mantener la gestión informada sobre el estado de sus negocios. Otras herramientas de BI son usadas para almacenar y analizar datos, tales como la minería de datos y el datawarehouse; sistemas de soporte o apoyo a las deci­siones y previsiones; almacén de documentos y gestión de documentos; gestión del conocimiento; mapeamiento, visualización de información y paneles (dashboards); sistemas de información de gestión, sistemas de información geográficas; análisis de tendencias; Software como servicio (SaaS) y otros. (Malhotra, 2001)

1. Análisis Avanzado: Análisis Avanzado: conocido como minería de datos y análisis predictivos, toma las ventajas de las técnicas de análisis estadís­ticos para predecir o proveer medidas de certeza sobre ciertos hechos. La gestión sobre el rendimiento de una organización (Portales, cuadros de mando, paneles de control): esta categoría general normalmente provee un sistema de varios com­ponentes interconectados, de tal forma que en conjunto describan una historia. Por ejemplo, un cuadro de mando integral que muestre componentes de indica­dores financieros, todos ellos combinados pueden describir métricas y patrones de aprendizaje y crecimiento en las organizaciones. (Gangadharan y Swami, 2004)
2. BI en tiempo real: permite la distribución de métricas en tiempo

real a través de emails, sistemas de mensajería instantánea y/o pantallas interac­tivas.

1. Datawarehouse y Datamarts: Datawarehouse y Datamarts: El datawarehouse es un componente importante de BI. El datawarehouse soporta la propagación física de los datos manejando grandes volúmenes de registros de las or­ganizaciones para integración, limpieza, agregación y tareas de consulta. (Gangadharan y Swami, 2004)

También puede contener datos operacionales, los cuales pueden ser definidos como un conjunto actualizable de datos integrados disponibles para toda una organización, para la toma de decisiones prácticas de un asunto especifico. Contiene datos vivos actualizados en tiempo real, no solamente fotos de un momento espe­cífico, y también conserva un historial mínimo. Las fuentes de datos pueden ser bases de datos operacionales, datos históricos, datos externos, por ejemplo, de or­ganizaciones de investigación de mercados o desde internet mismo, o información desde un entorno de datawarehouse ya existente. Las fuentes de datos pueden ser de bases de datos relacionales o cualquier otra estructura que apoya o soporta el conjunto de sistemas transaccionales de una organización. (Gangadharan y Swami, 2004)

Un datamart, tal como se describe en (Kumari, 2013), es una colección de disciplinas organizadas para el apoyo de las decisiones basadas en las necesidades de un departamento dado. Finanzas tiene su datamarts, marketing tiene el suyo, y ventas tienen la suya y así sucesivamente.

Cada departamento tiene su propia interpretación de cómo debe verse un datamart y el datamart de cada departamento es particular y atiende las necesida­des específicas del área. Similar al datawarehouse, los datamarts contienen datos transaccionales que ayudan a expertos en negocios a crear una estrategia basada en el análisis de las tendencias y experiencias pasadas. La principal diferencia es que la creación de los datamarts se basa en una necesidad específica, predefinida para un grupo determinado. Un datamart puede apoyar o soportar procesos o unidades de negocio específicos (Ranjan, 2009).

Las herramientas de BI son ampliamente aceptadas como una capa intermedia entre aplicaciones transaccionales y aplicaciones de apoyo a la toma de decisión, estas están desacopladas y extraen informaciones de transacciones de negocio. Las habilidades de BI incluyen, apoyo a la decisión, procesamiento analítico en línea, análisis estadísticos, análisis predictivo, y la minería de datos (Ranjan, 2009).

Las fuentes de datos pueden ser bases de datos operacionales, datos históricos, datos externos, por ejemplo, desde las empresas de investigación de mercados o desde internet, datos no estructurados de redes sociales, o información desde un datawarehouse existente. Las fuentes de datos pueden ser bases de datos relaciona­les o cualquier otra estructura de datos de sistemas transaccionales. Ellos también pueden residir en muchas plataformas diferentes, tales como tablas, hojas de cálculo, o información no estructurada, tales como archivos de texto plano o imágenes y otras informaciones multimedia (Ranjan, 2009).

1. Beneficios de BI

BI provee muchos beneficios a las compartías que lo utilizan. Puede ayudar a eliminar muchas conjeturas erróneas dentro de una organización, mejorando la comunicación entre departamentos mientras se coordinan las actividades, y apoyan a las organizaciones para responder rápidamente a cambios de condiciones finan­cieras o preferencias de clientes. BI puede ayudar a mejorar el rendimiento general de una organización (Ranjan, 2009).

La información es frecuentemente considerada como el segundo recurso mas

importante que una compañía tiene (lo más valorable de una compañía son las personas). Cuando una compañía puede tomar decisiones basadas en información oportuna y precisa, puede ayudar a mejorar su rendimiento en su segmento de mercado. BI también agiliza la toma de decisiones, ayuda a actuar rápida y correc­tamente con la información adecuada antes de otras empresas de la competencia. También pueden mejorar la experiencia del cliente, teniendo en cuenta la respuesta oportuna y adecuada a los problemas y prioridades de los mismos. A continuation se listan algunos de estos beneficios:

* Con herramientas de BI, los empleados pueden fácilmente convertir sus cono­cimientos de negocio en inteligencia analítica para resolver muchas cuestiones de negocio, tales como incrementar la tasa de respuestas desde correos electrónicos, teléfonos, y mejorar las campañas de ventas desde internet.
* Con BI, las empresas pueden identificar sus clientes más rentables y las razo­nes subyacentes para la lealtad de esos clientes, así como identificar clientes futuros con grandes potenciales.
* Analizar los datos de clics para mejorar las estrategias de comercio electró­nico.
* Detección rápidamente de problemas reportados de productos para minimi­zar el impacto de las deficiencias en sus diseños.
* Descubrir lavado de dinero de actividades delictivas.
* Analizar la rentabilidad potencial del cliente, y reducir el riesgo a través de una puntuación más precisa de crédito financiero de los mismos.
* Determinar cuáles son las combinaciones de productos y servicios que los clientes son más propensos a comprar y cuando.
* Analizar los ensayos clínicos de fármacos experimentales.
* Establecer tarifas más rentables para las primas de seguros.
* Reducir el tiempo fuera de un equipamiento mediante la aplicación de man­tenimiento predictivo.
* Determinar con el análisis de deserción y rotación de clientes, la causa por la cual los clientes se van a los competidores o se convierten en nuestros clientes.
* Detectar y disuadir comportamientos fraudulentos, por ejemplo, de picos de uso cuando las tarjetas de crédito o tarjetas telefónicas son robadas.
* Identificar nuevos compuestos de fármacos moleculares prometedores.

1. Tecnología de BI

La inteligencia empresarial provee datos organizacionales de tal manera que los filtros de conocimientos organizacionales puedan fácilmente asociarse con estos datos y volverlos en información para la organización. Las personas involucradas en procesos de inteligencia de negocios podrían usar software y otras tecnologías para reunir, almacenar, analizar, proveer accesos a datos, y presentar esos datos de una manera simple y útil.

El software ayuda en la gestión de una organización, y a las personas a hacer mejores decisiones de negocios, teniendo la información precisa, actualizada, y re­levante cuando lo necesiten. Algunas empresas usan datawarehouse porque es un conjunto de información lógica recolectado desde varias bases de datos operacionales con el objetivo de crear inteligencia de negocios (Ranjan, 2009).

Para que los sistemas BI trabajen efectivamente, existen algunas restricciones técnicas que deberían ser tratadas:

* Seguridad y acceso de usuarios al datawarehouse.
* Volumen de datos (capacidad).
* Cuanto tiempo será almacenado el dato (retención de datos).
* Sizing y rendimiento de infraestructura (servidores).

Las personas que trabajan en BI desarrollan productos que facilitan el traba­jo, especialmente cuando las tareas de inteligencia involucran conseguir y analizar grandes cantidades de datos no estructurados. Cada proveedor tópicamente define BI de una forma particular, y comercializa herramientas para hacer BI de la for­ma en que cada uno lo propone. BI incluye herramientas en diversas categorías, incluyendo las siguientes: (Ranjan, 2009).

* AQL (Associative Query Logic) - Lógica Asociativa de Consultas.
* Métricas y mediciones del rendimiento del negocio.
* Planeamiento Empresarial.
* Data mining (DM), Data Farming, y Datawarehouses.
* Sistemas de apoyo a la decisión (DSS) y predicción.
* Datawarehouse de documentos y gestión documental.
* Sistema de Gestión Empresarial.
* Finanzas y presupuestos.
* Recursos humanos.
* Gestión del conocimiento.
* Mapeamiento, visualización de la información, y paneles de control (dash­boards).
* Sistemas de gestión de informaciones.
* Sistemas de información geográfica (GIS).
* OLAP (Online Analytical Processing) y análisis multidimensional; a veces simplemente llamado “Analytics “(basado también en “hipercubo “o “cubo “).
* BI en tiempo real.
* Análisis de datos estadísticos y técnicos.
* Gestión de la línea de producción, Gestión de demandas.
* Gestión de la cadena de Suministro/Gestión de la cadena de demanda.
* Análisis de tendencias.
* Reportes y consultas de usuarios/usuarios-finales.

BI frecuentemente usa indicadores de rendimientos (KPIs, key performance in-

dicators) para evaluar el estado actual de los negocios y para establecer un plan de acción. Más y más organizaciones han comenzado a disponibilizar más datos con mayor velocidad. En el pasado, los datos solo estaban disponibles después de uno o dos meses, lo que no ayudaba a los directivos de empresas para ajustar las actividades con la velocidad necesaria para alcanzar sus objetivos. Recientemen­te, los bancos han intentado disponibilizar los datos en el intervalo más corto y reduciendo los atrasos (Ranjan, 2009).

Por ejemplo, para negocios de alto riesgo operacional (por ejemplo, tarjetas de créditos), un banco multinacional disponibiliza los datos relacionados con KPI semanalmente, y en ocasiones ofrece un análisis diario de los números. Esto signi­fica que los datos normalmente están disponibles a cada 24 horas, requiriendo la automatización y el uso de sistemas de TI.

1. Breve discusión

La experiencia actual de cualquier nueva forma de organización es la cadena de valor, la cual es un conjunto de actividades primarias y secundarias que crea valor para los clientes. (Denison, 1997) examina muchas actividades críticas relacionada a la cadena de valor. Sin un BI eficaz para dirigir las organizaciones orientadas a los procesos de apoyo, esto no sería posible.

(Davenport, 1993) describe varias cuestiones en la reingeniería en las innova­ciones de los procesos de negocio. De acuerdo a (Adelman, Moss, y Barbusinski, 2002), BI es un término que engloba un amplio rango de software de análisis y soluciones para recolectar, consolidar, analizar y proveer acceso a la información de una manera sencilla para que los usuarios de una empresa puedan tomar me­jores decisiones de negocio. (Malhotra, 2001) describe a BI como un facilitador de conexiones en una nueva forma de organización, trayendo información en tiempo real para centralizar repositorios y apoyar el análisis, que puede ser explotada en cada nivel horizontal y vertical, dentro y fuera de la empresa.

Bi describe el resultado de un análisis profundo de los datos detallados del negocio, incluyendo base de datos y tecnologías de aplicación, asá como prácticas de análisis (Gangadharan y Swami, 2004). BI es técnicamente más amplio, lo que potencialmente engloba la gestión del conocimiento, la planificación de recursos empresariales, sistemas de apoyo a la toma de decisiones y la minería de datos (Gangadharan y Swami, 2004).

(Nguyen, Schiefer, y Tjoa, 2005) introdujeron una arquitectura mejorada de BI que cubre un proceso completo para identificar, interpretar, predecir, automatizar y responder a los ambientes de negocios; y por lo tanto tiene como objetivo reducir el tiempo de reacción necesario para las decisiones empresariales. (Nguyen y cols., 2005) propone una infraestructura de TI basada en eventos para operar aplicacio­nes de BI que permiten análisis en tiempo real a través de procesos de negocios corporativos, y brindar recomendaciones automáticamente para optimizar las operaciones comerciales, y cerrando efectivamente la brecha entre sistemas de BI y procesos de negocio. (Seufert y Schiefer, 2005) sugieren una arquitectura de BI mejorada, que tiene como objetivo aumentar el valor de la información mediante la reducción del tiempo de acción y la interconexión de los procesos de negocio en la toma de decisiones. Las empresas no solo desean conocer lo que ha sucedido, sino necesitan saber las causas subyacentes. Por ejemplo, en lugar de saber cuántas mantas fueron vendidas en un mes, las empresas desean entender cuántas fueron vendidas en un país determinado durante un evento meteorológico. BI proporcio­na una visión integrada unificada de las actividades empresariales. Las empresas han construido sistemas de BI que apoyan análisis de negocio y de toma de de­cisiones para ayudarlos a un mejor entendimiento de sus operaciones y competir en el mercado (Gangadharan y Swami, 2004). Algunas innovaciones en tecnolo­gías de almacenamiento de datos están superando significativamente el progreso en potencia de procesamiento Unidad Central de Procesos (CPU), anunciando una nueva era para BI en tiempo real. Como resultado, algunos proveedores de soft­ware con herramientas superiores ofrecen una suite completa de aplicaciones para análisis de BI, herramientas y modelos de datos que permiten a una organización aprovechar su información. Las herramientas BI facilitan el acceso a un gran volu­men de datos corporativos, y convertir esos datos en información útil y procesable que sea consistente a través de la versión coherente de la verdad. Las empresas aún sienten que BI tiene complejidades relacionadas con la tecnología y que pue­de usarse solamente por especialistas con conocimientos técnicos, además que los costos de implantación son altos. Las empresas requieren estos análisis en tiempo real para los proyectos a corto plazo. El BI tradicional puede que no haga esto, pero en un ambiente BI en tiempo real ciertamente podría atender las necesida­des actuales de las empresas. Los datos finalmente son considerados como recursos corporativos en una nueva disciplina. Cualquier sistema transaccional (incluyen­do Sistemas de Planificación de Recursos Empresariales (ERP) y Administración basada en la relación con los clientes (CRM)) y cualquier aplicación de apoyo a la decisión (incluyendo datawarehouses y datamarts) son BI, si y solo si fueron desarrollados bajo la protección y la metodología de una iniciativa estratégica de toda la organización (Gangadharan y Swami, 2004). Los sistemas tradicionales de BI consisten en una base de datos en el backend, una interfaz de usuario en el frontend, software que procesa la información para producir la propia inteligen­cia de negocios, y un sistema de informes. Las capacidades de BI incluyen apoyo a la decisión, procesamiento analítico en línea, análisis estadísticos, predicción y minería de datos. Diferentes sectores como fabricantes, comercios electrónicos, em­presas de telecomunicaciones, aerolíneas, minoristas, sistemas de salud, servicios financieros, bioinformática y hoteles utilizan BI para apoyo a clientes, investigación de mercado, segmentación, rentabilidad del producto, análisis y distribución

de stock, análisis estadístico, informes multidimensional, detección fraudes, en­tre otros. BI y minería de datos es un área que está fuertemente influenciado por técnicas estadísticas tradicionales, y la mayoría de los métodos de minería de datos revela una fuerte base de métodos estadísticos y de análisis de datos. Algunas de las técnicas tradicionales de minería de datos incluyen clasificación, agrupación, análisis de valores atípicos, patrones secuenciales, análisis de series temporales, la predicción, la regresión, análisis de enlaces (asociaciones), y métodos multidimen­sional incluyendo el procesamiento analítico en línea Procesamiento analítico en línea (OLAP). Estos pueden clasificarse en una serie de técnicas de minería de datos, que se clasifican e ilustran en la Tabla 1 (Goebel y Gruenwald, 1999).

|  |  |
| --- | --- |
| TECNICAS | DESCRIPCION |
| Modelo predictivo | Predecir valor para un atributo especí­fico del elemento de datos. |
| Caracterización y minería de datos des­criptivo | Distribución, dispersión y excepción de datos |
| Asociación, correlación, análisis de la causalidad (Análisis Link) | Identificar relation entre atributos |
| Clasificación | Determinar a qué clase pertenece un elemento de datos |
| La agrupación y análisis de valores atí­picos | Partición de un conjunto en clases, con lo cual elementos con características si­milares se agrupan |
| Análisis de patrones temporal y se- cuencial | Tendencia y desviación, patrones se­cuenciales, frecuencia |
| OLAP (Procesamiento Analítico en Línea) | Herramientas OLAP permiten a los usuarios analizar distintas dimensiones de datos multidimensionales. Por ejem­plo, proporciona series temporales y puntos de vista de análisis de tenden­cias. |
| Modelo de visualization | Hacer fácil la descubierta de cono­cimiento usando charts, plots, histo­grams y otros medios visuales |
| Análisis Exploratorio de Datos(EDA) | Explorar un conjunto de datos sin una fuerte dependencia en hipótesis o mo­delos; el objetivo es identificar patrones de una manera exploratoria |

Cuadro 2.1: Técnicas actuales de BI. (Goebel y Gruenwald, 1999)

En el siguiente capítulo se presenta una introducción a los métodos y técnicas de análisis exploratorio, y en especial la técnica actualmente llamada Data Discovery, la cual fue aplicada en este trabajo.

1. Data Discovery Analysis

Data discovery es una arquitectura de BI destinado a informes interactivos y en tiempo real que pueden ser explorados desde múltiples orígenes (Marakas, 2003). La mayor parte de la base instalada en todo BI son propietarias de empresas tra­dicionales que han construido sus plataformas alrededor de una capa semántica y metadatos, la cual es generalmente accesible solo por herramientas del propio fabricante. La situación actual de carácter propietario de la capa semántica tradi­cional de BI fue aceptada y adoptada por mas organizaciones como un facilitador para análisis ad hoc a cambio de una única y confiable versión de la realidad, que puede ser accedida fácilmente por los usuarios de negocio, ocultando los aspectos técnicos y la complejidad de las estructuras de datos subyacentes. Sin embargo, con la adopción y rápido crecimiento de las herramientas de data discovery como Qlik, Tableau, y Tibco Spotfire, los usuarios buscan cada vez más acceso a datos confiables en capas semánticas cerradas, y los proveedores de BI se enfrentan a un reto difícil para seguir siendo relevantes en un mercado que está en transición. La respuesta de los proveedores tradicionales y la inversión significativa hasta la fecha, ha sido la utilización de sus capas semánticas existentes para promover la gestión a nivel empresarial de sus propias herramientas de data discovery desarrolladas internamente. Esto ha sido un mecanismo para diferenciar sus soluciones de las de proveedores de data discovery pure-play. Mientras que, en teoría, este enfoque logra un equilibrio entre la facilidad de uso y escalabilidad empresarial. Esto ha mostrado poco éxito para la mayoría de los proveedores de BI tradicionales como huecos de importantes funcionalidades que permanecen entre sus herramientas de data discovery desarrolladas internamente y los de los proveedores especialistas de data discovery. (Josh Parenteau, 2015)

La capa semántica que sirve como la base de la mayoría de las plataformas BI tradicionales ha sido ampliamente adoptada por muchas organizaciones a través de los años y ha sido promovida y generalmente aceptada como un componen­te esencial de una plataforma BI. Proveedores como SAP (BusinessOjects), IBM (Cognos) u Oracle (OBIEE) mantienen una gran base instalada de clientes que han invertido mucho en el desarrollo, operación y mejora de estas plataformas construidas alrededor de una definida y centralizada capa semántica propietaria.

Este enfoque ha funcionado bien cuando el objetivo fuera una única fuente de datos, en apoyo a sistemas definidos de registros centralizados de informes y gestión de dashboards fomentando la consistencia, la gobernanza e integración entre las plataformas de presentación y capas de metadatos. Sin embargo, con el surgimiento y expansión de data discovery, el concepto de auto servicio cobro preponderancia.

Los usuarios de negocio y análisis ahora tienen acceso a un gran rango de herra­mientas que promueven y apoyan el uso autónomo sin la participación de TI. Como tal, hay una necesidad emergente para acceder a las reglas de negocio integradas dentro de la capa semántica propietaria de herramientas de BI existentes.

Mientras esto no es posible aun en la mayoría de los casos hoy en día, algunos proveedores ya han comenzado a adoptar un enfoque cada vez más abierto para sus capas semánticas propietarias, y esto puede llevar a un cambio mayor de mer­cado con el tiempo. La oportunidad probablemente será dictada por el éxito que los proveedores tradicionales de BI tengan con el desarrollo y la adopción de sus propias ofertas de data discovery, que se ha limitado hasta la fecha.

El acceso abierto a metadatos no es inédito en BI y en mercados analíticos. Ejemplos incluyen Oracle OBI EE, Microsoft Power, SAP BusinessObjects y, más recientemente, conectividad nativa de Tableau para modelo de datos Birst a través de su capa semántica propietaria.

Oracle BI Enterprise Edition (OBI EE) fue uno de los primeros productos de plataformas BI para tener una capa semántica abierta, un vestigio de los orígenes del producto como una nueva plataforma web abierta, desarrollada por nQuire, posteriormente adquirida por Siebel en 2001, y finalmente por Oracle en el 2005. Con OBI EE, el modelo de datos puede ser publicado y accedido con una conexión ODBC que puede ser consumida por herramientas e interfaces de terceros. Inicial­mente, esto es cómo Oracle proporciona conectividad a su herramienta interna de informes de producción desarrollada, Oracle BI Publisher. Más allá de eso, pocos clientes son conscientes de esta capacidad, y citan los malos resultados como una razón por la que no fue adoptado ampliamente.

La reciente alianza entre Birst y Tableau, establecida en abril 2015, es el más reciente ejemplo de este cambio hacia acceso abierto e integración entre proveedores puros de data discovery y fabricantes tradicionales de BI. Antes del anuncio, Birst no había permitido el acceso a su estructura de datos propietaria o capa semántica, hizo accesible sólo a de informes, dashboard y capacidades data discovery auto- contenidas dentro de la cartera Birst. A través de la alianza y desarrollo conjunto, se añadió una conexión con Tableau que permite la conexión directa con el medio ambiente Birst. Esto proporciona una mayor flexibilidad y una gama más amplia de opciones a los clientes comunes.

Una alianza similar a la que existe entre Birst y Tableau fue anunciada en 2014 entre SAP y Microsoft permite a los usuarios Power Query acceder a la capa semántica de SAP BusinessObjects (Universe). Este acuerdo permite a los clientes comunes la opción de usar herramientas de Microsoft Power BI para data discovery a la vez aprovechan las inversiones en el SAP BusinessObjects Universe.

Un inconveniente de estas alianzas es que son actualmente unidireccionales en su naturaleza y solo otorgan acceso de sólo lectura a herramientas de data discovery

de otros proveedores de BI a la capa semántica propietaria de proveedores de BI tradicionales. Como tal, ellos todavía no apoyan la promoción de modelos de datos derivados de herramientas de data discovery en la capa semántica como una forma de promover a proveedores independientes que rigen capacidades de data discovery. Esto es, sin embargo, una oportunidad potencial de modernización que los proveedores tradicionales pueden considerar.

Proveedores de software independientes e integradores de sistemas desarrollaran nuevas soluciones de capa intermedia que facilitan el acceso a la capa semántica a las herramientas de data discovery a través de servicios web.

Mientras que los clientes prefieren una única solución de sus proveedores titu­lares, ya sea de los proveedores de plataforma tradicional de BI o de proveedor de data discovery, los clientes actualmente encontrarán más oportunidades para acceder a la capa semántica desde los integradores de software y los proveedores independientes menores.

1. Framework de Análisis de Negocio de Gartner

En este trabajo fue adoptado un Framework de Análisis de Negocio de Gartner, el cual se describe a continuación.

Hay una serie de defectos relacionados en la mayoría de las organizaciones con relación a BI y plataformas analíticas, asá como una percepción errónea de sus objetivos y como gestionarlos. Los equipos de BI, especialmente si se encuentran en el área de TI, creen que:

* La plataforma analítica de negocio debe ser una solución estrechamente inte­grada con pocos componentes, preferentemente de un único proveedor, para entregar una sola versión de la verdad para la organización.
* La información puede ser confiable sólo si está almacenada en un datawarehouse corporativo y entregada a los consumidores de información usando artefactos de BI, tales como informes y dashboards.
* Las informaciones creadas o manipuladas por los usuarios de negocios inevi­tablemente producirán discrepancias a través de diferentes análisis, lo que lleva a decisiones equivocadas, generando caos en la organización a través del tiempo.
* La responsabilidad del departamento TI para la gestión de información ter­mina en la capa semántica de BI y en los contenidos orientados a TI.
* Procesos analíticos orientados al negocio están fuera del alcance y no soporta­dos por TI. Hay varios problemas orientados a la información documentados en el mundo BI y análisis, que obligan a los líderes de BI a seguir estas creencias y desplegar un entorno de BI centralizado y monolítico, que termi­na siendo impuesta a los usuarios, independientemente de su adecuación a las necesidades.
* Los proveedores licenciados de BI, a favor de sus propias plataformas, apoyan

este enfoque.

Según Gartner, una plataforma analítica necesita evolucionar más allá del pensamiento monolítico. Debe ocurrir una transformación para ofrecer dife­rentes soluciones para las diferentes necesidades del usuario, con un conjunto diverso de niveles de integración, y encontrar un equilibrio entre confianza y agilidad. El propósito es ayudar a los usuarios a alcanzar sus objetivos de negocio a través del uso de la tecnología apropiada, no para erradicar las solu­ciones de BI orientados al usuario que resuelven parcialmente sus problemas de hoy.

El entorno de BI y el análisis resultante también exigirá cambios en los procesos de gestión de la información, como atribuir nuevas responsabilidades a diferentes personas en la organización.

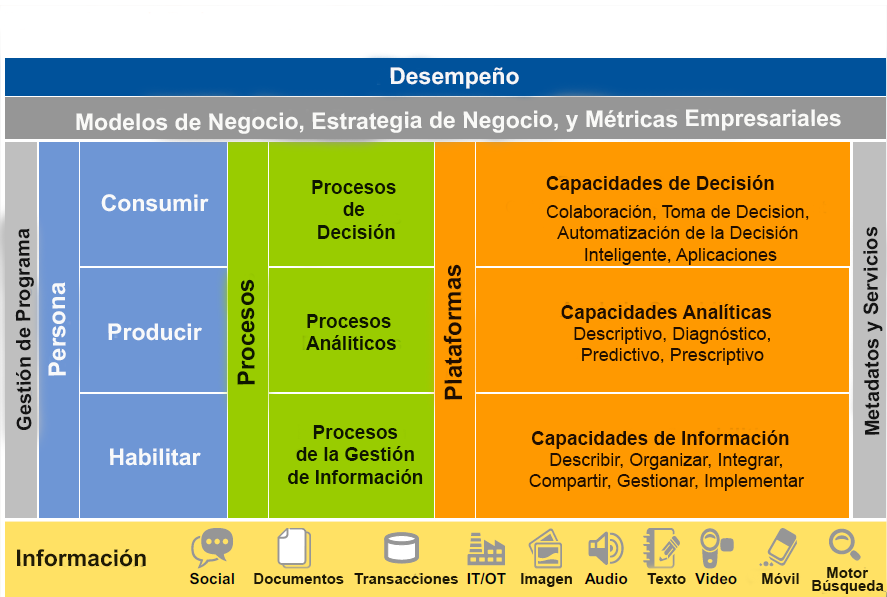
A continuación, el framework de análisis de negocios. La figura 2.2 presenta el framework de análisis de negocios de Gartner.

Figura 2.2: Framework Analítico de Negocios Gartner. Tapadinhas, J. (2014)

El framework de análisis de Gartner identifica las personas, los procesos y com-

ponentes de la plataforma que apoyan la transformación de la información en un mejor rendimiento de la organización. El uso de esta herramienta está hecho por la lectura desde arriba hacia abajo, comenzando con los resultados del negocio y luego descifrando las composiciones analíticas de apoyo y la información necesaria para alcanzarlos. De acuerdo a las necesidades de los usuarios, la plataforma debe ser rediseñada con un amplio conjunto de capacidades técnicas (llenando los va­cíos), nuevas responsabilidades y organización. Centrándose en las herramientas o normalización de proveedor por sí sola no es la respuesta.

El framework es también muy útil para definir los estados actuales y futuros de la arquitectura. La diferencia entre ellos es el mapa de rutas e incluye cambios en personas y procesos. La organización muy probablemente también necesitara re-organizar y capacitar a los proveedores y usuarios de BI y análisis. Los usuarios de negocio deben ganar acceso a las herramientas analíticas adecuadas, de acuerdo a sus metas y habilidades, y un rango comprensivo de fuente de datos con tipos de datos variados, granularidad adecuada y accesos apropiados. Teniendo en cuenta el espectro de capacidades analíticas con énfasis en las plataformas, en particular, el componente de capacidad analítica, podemos notar cuatro estilos analíticos que se detallaran a continuación (Figura 2.3).

Análisis

Entrada Humana

Figura 2.3: Espectro Analítico. Tapadinhas, J. (2014)

Las capacidades analíticas implementadas en organizaciones son a menudo li­mitadas el análisis descriptivo, a través de reportes básicos y dashboards. Con esto, la pregunta, ¿“Que paso? “puede ser respondida. Después de conocer “Que, “,lo más probable es que los usuarios también pregunten, “Por qué paso? “. Abordar adecuadamente esto requiere mucha más agilidad y más capacidades avanzadas de exploración de la información. Despliegues de BI tradicionales tienden a tener huecos en esta área, pero TI por lo general pasa por alto el impacto de esta problemática y continúa impulsando el estándar del proveedor y sus herramientas no aptas-para-propósito. Como consecuencia, los usuarios recurrirán a Excel, consul­tas ad hoc, extracciones de datos y a los equipos de shadow TI para lograr sus metas de análisis.

Los líderes de BI deben extender el BI y la plataforma analítica hasta el análisis de diagnóstico para complementar el análisis descriptivo. Aquí es donde OLAP y los modelos de datos en memoria son utilizadas para proporcionar una navegación fácil y rápida de datos sin una consulta predefinida. Aprovechando mejoras en el nivel de acceso a datos, también vemos la necesidad de mejorar las capas semánticas abstrayendo la complejidad del modelo físico subyacente. Esto puede hacer que sea mucho más fácil para el descubrimiento de autoservicio sin el cuello de botella de TI que se encuentra en un típico equipo de BI.

Más allá de la capa de datos, vemos la introducción de nuevas herramientas de visualizaciones de datos, y aquí es donde el enfoque del rápido crecimiento de las herramientas de data discovery se concentran. Pero herramientas tradicionales pueden también proporcionar mejoras con un mayor enfoque en informes más comprensivos (como el análisis de varianza), planificación integrada, dashboards e informes KPI.

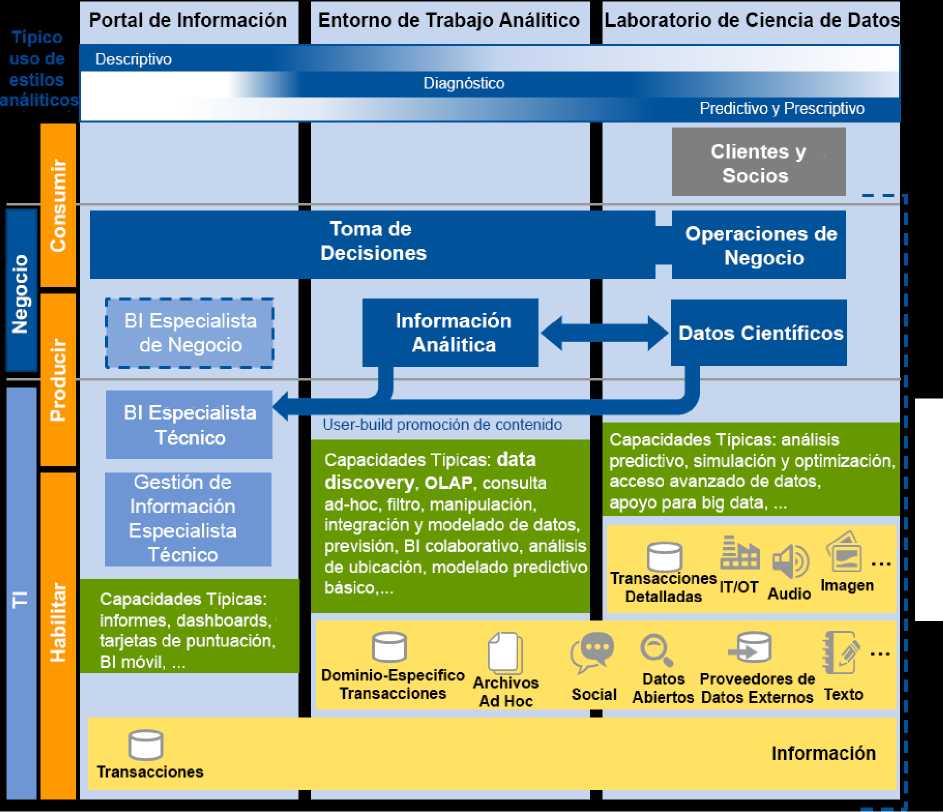
A través del tiempo, con un crecimiento a un alto nivel de madurez de aná­lisis, la organización debería moverse dentro del análisis predictivo y preceptivo. Esto requiere un incremento significante en los niveles de habilidades del analista de negocios. Modelos predictivos requieren desarrollo y mantenimiento con lógicas complejas y reglas de negocio. Ellos incorporan métodos sofisticados que pueden también requerir un entendimiento profundo de la estadística o investigación operacional.

Además, las organizaciones deben darse cuenta de que hay necesidad de mezclar todas estas diferentes técnicas en soluciones integrales en lugar de dejarlos aislados.

Rediseñar el BI y la Plataforma Analítica Los líderes de BI deben seguir las herramientas fundamentales descritas anteriormente para asá con éxito rediseñar el BI y la plataforma de análisis. Gartner recomienda la instalación de una arqui­tectura por niveles compuesta por:

* Portal de Información.
* Workbench Analítico.
* Laboratorio de datos científicos.

En la figura 2.4 se presenta la representación de BI en niveles y la plataforma

analítica, la cual puede ser utilizada como una guía genérica que puede ser ajustado de acuerdo a las características específicas de la organización.

**Gestión - Metadatos**

Para hacer realidad la visión de los tres niveles y ser capaz de maximizar sus fortalezas, los líderes de BI necesitan implementar nuevas capacidades técnicas para proporcionar nuevos estilos de análisis, mejorar el uso de las herramientas existentes a través de una mejor integración global, y proporcionar metadatos comunes y gobernanza. Procesos, roles de personas y responsabilidades, son de suma importancia para el éxito.

Figura 2.4: Típico uso de estilos analíticos. Tapadinhas, J. (2014)

Ellos deben ser tratados en conjunto con la plataforma técnica como se describe

en el framework de análisis de negocios de Gartner.

Vamos a ampliar cada nivel para entender cómo integrar y aprovecharlos en conjunto.

1. Portal de información

Seguir de cerca las características de sistemas de registros desde el Pace Layer Model. El portal de información es el área de trabajo donde los usuarios de negocio pueden encontrar rápida y fácilmente las métricas clave de confianza con la cual la organización mide su rendimiento. Por lo general hecho de capacidades de informes y dashboard que proporciona contenido para los consumidores de información. Sus productos son el resultado de un proceso de desarrollo formal que abarca que un usuario de negocios establezca requisitos y un especialista técnico (típicamente de TI, pero cada vez más de negocio) los implemente.

Capacidades típicas de la plataforma:

* Informes.
* Dashboards.
* Integración Microsoft Office.
* BI móvil.
* Análisis integrado.

Ejemplo de herramientas y proveedores:

* SAP BusinessObjects.
* IBM Cognos.
* Oracle BI.
* Microsoft Reporting Services.
* MicroStrategy.
* Información Builders WebFocus.

1. Workbench analítico

El workbench analítico es el área de trabajo usado para investigar tendencias en indicadores de confianza o detectar patrones en otros conjuntos de datos — desde múltiples fuentes -. que pueden convertirse en oportunidades o riesgos. Es una capa ágil para explorar información y tener acceso a una amplia gama de fuentes de datos, con limitado o ningún apoyo de expertos técnicos. Los conjuntos de herramientas deberían incluir una herramienta de data discovery (descubrimiento de datos) y un número de otras capacidades para ayudar a los usuarios de negocios a extraer valor de la información de forma autónoma.

En el espectro de análisis, el workbench es capaz de proporcionar análisis des­criptivo, pero por lo general se centrará en el análisis de diagnóstico. En algunos casos - es decir, a través del uso de herramientas de data discovery más centrado al análisis. - Puede extender a un nivel básico de análisis predictivo y ganara el modelado de datos y capacidades analíticas más avanzadas en el futuro.

Capacidades típicas de la plataforma:

* Data discovery.
* Informes/consultas Ad hoc.
* Inteligencia geoespacial y localización.
* Análisis integrado avanzado.

- OLAP.

* Mashup de datos y modelado de usuarios de negocios.
* Colaboración.
* Filtrado y manipulación de datos.

Ejemplo de herramientas y proveedores:

* Tableau Software.
* Qlik.
* Tibco Spotfire.
* SAS Visual Analytics.
* SAP Lumira.
* Oracle Endeca Información Discovery.
* MicroStrategy Visual Insight.
* Alteryx.
* Microsoft SQL Server Analysis Services and Power BI.

1. Laboratorio de datos científicos

El laboratorio científico de datos es el área de trabajo donde análisis avanzados se llevan a cabo y es la incubadora ideal para iniciativas big data. Es un entorno flexible donde experimentos de prueba y error es actualmente alentado para generar ideas impactantes para la organización.

Un amplio conjunto de capacidades técnicas es esperado y a menudo propor­cionado por herramientas especializadas con integración TI mínima, destinada a entregar agilidad y capacidad de responder las preguntas imprevisibles. Esto es porque TI tiende a pasar por alto esta área a favor de inversión en el portal de información. Los usuarios son cualificados y experimentados, a menudo más que los expertos técnicos en TI. Sus conjuntos de herramientas incluyen capacidades de data mining, predicciones y otras herramientas de estadísticas y análisis complejos.

Capacidades típicas de la plataforma:

* Acceso avanzado de datos.
* Soporte para fuentes de big data.
* Análisis descriptivo avanzado.
* Análisis predictivo.
* Prediction.
* Optimization.
* Simulation.
* Otros análisis avanzados.
* Aunque no capacidades BI, Hadoop y bases de datos NoSQL también deben ser referenciados aquí.

Ejemplo de herramientas y proveedores:

* SAS Enterprise Miner.
* IBM SPSS.
* SAP InfiniteInsight.
* Revolution Analytics y R.
* RapidMiner.
* Knime.
* Alteryx.
* FICO.
* Dell StatSoft.
* Cloudera.
* Hortonworks.
* MapR y otras distribuciones Hadoop.

Entender las características de BI en niveles y la plataforma analítica. La si­guiente tabla resume las características que posee cada una de las tres capas. Los líderes de BI deberían deberán intentar entender los huecos en su plataforma ana­lítica, y cambiar sus estrategias en consecuencia.



Figura 2.5: Características de BI en niveles y Plataforma Analítica. Tapadinhas, J. (2014)

CAPITULO 3

1. Selección de la herramienta para Data Dis­covery

Fueron analizados los estudios de Gartner (Herschel, Linden, y Kart, 2015) (Sallam, 2015) para la selección de la mejor herramienta que se adecue a los criterios nece­sarios para ser utilizado en esta tesis. Estos documentos realizan un análisis de las mejores herramientas del mercado, en un área de conocimiento. A continuación, se presenta el Cuadrante Mágico de Gartner, para herramientas de BI y Analytics.

1. Cuadrante Mágico de Gartner

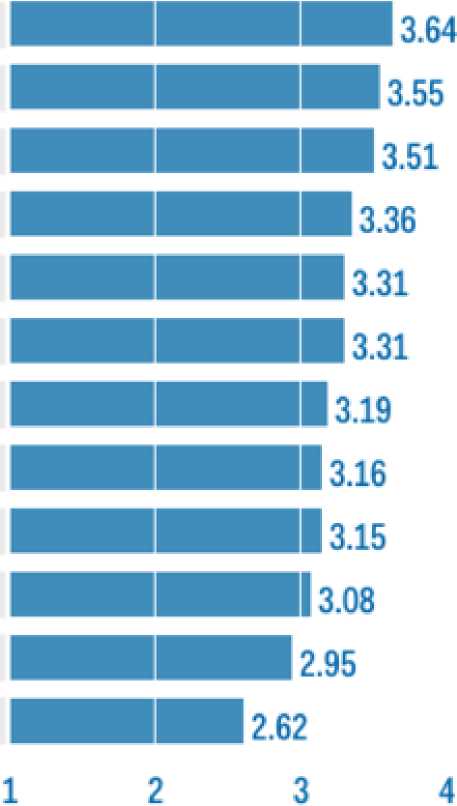
Figura 3.1: Cuadrante Mágico para BI y Plataformas Analíticas. Herschel, G., Linden, A., y Kart, L. (2015)

Los líderes del mercado se encuentran siempre en el cuadrante superior derecho. Se puede observar una amplia diferencia entre Tableau y los demás líderes.

**CAPACIDAD PARA EJECUTAR**

En las siguientes figuras se presenta el análisis de Gartner que evalúa las capa­cidades críticas que debe tener una herramienta de BI y Analytics, para adecuarse a las necesidades del mercado.

Product or Service Scores for Decentralized Analytics

Tibco Tableau SAS Institute Qlik

MicroStrategy

Birst

IBM

Log ¡Analytics SAP

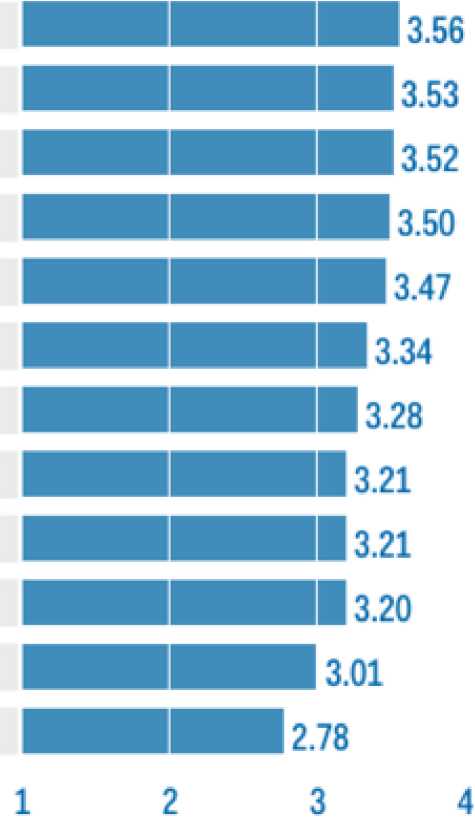
information Builders

Microsoft

Oracle

Figura 3.2: Puntuaciones de producto o servicio para análisis descentralizado. Sallam, H. S. O. T. H., Parenteau. (2015)

Product or Service Scores for Governed Data Discovery

Tableau

SAS Institute

MicroStrategy

Tibco

Birst

QEk

IBM

Log ¡Analytics Information Builders SAP

Microsoft

Oracle

Figura 3.3: Puntuaciones de producto o servicio guiados por Data Discovery. Sallam, H. S. O. T. H., Parenteau. (2015)

Los valores posibles van del 1 al 5, conforme la siguiente evaluación:

1. Pobre o ausente: la mayoría de los requisitos de esta capacidad no fueron alcanzadas.
2. Justo: Algunos de los requisitos fueron alcanzados.
3. Bueno: cumple con los requisitos.
4. Excelente: alcanza o excede algunos requisitos.
5. Superior: excede significativamente los requisitos.

Tableau tiene una posición fuerte en capacidad de ejecución (producto/servicio, su oferta, ejecución de ventas, marketing, experiencia del cliente) en el eje de líderes del cuadrante. Esta herramienta fue la que mejor se adecuo a las necesidades del trabajo de Tesis, dado que cuenta con una versión publica para la construcción y publicación de dashboards, además de la facilidad de uso que nos proporciona. Tableau Desktop, la cual se basa en tecnología drag and drop (arrastrar y soltar) permite analizar datos rápidamente y permite ver los cambios en tiempo real sin necesidad de codificación, de esta manera, posibilita a un usuario con escasos conocimientos técnicos, poder utilizarlo de igual manera.

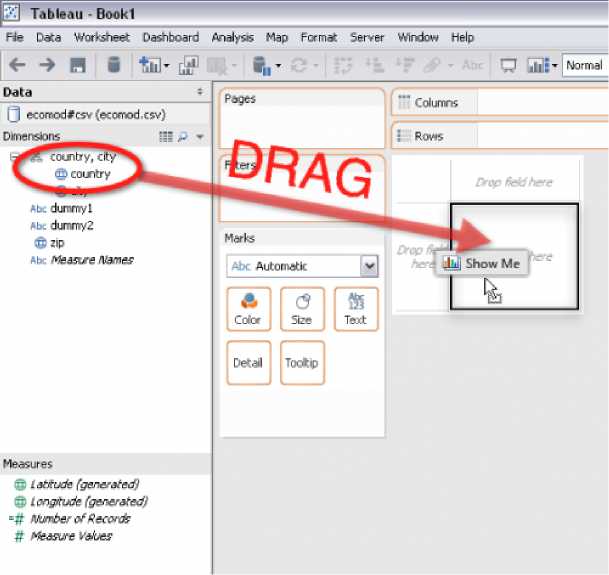


Figura 3.4: Arrastre el campo país para el campo desplegable señalado. Peck, G. (2013).

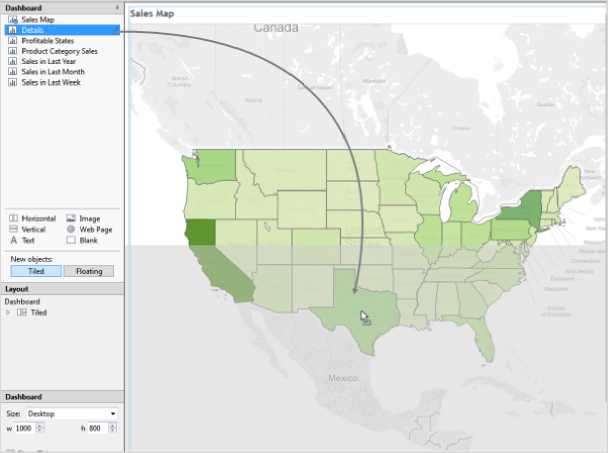


Figura 3.5: Arrastrar hojas de trabajo al dashboard. Peck, G. (2013).

De una forma ágil el usuario puede conectarse a diversas fuentes de datos y crear paneles interactivos, conectando entre sí los diferentes componentes (tipo de gráfico) que proporciona la herramienta. La herramienta permite utilizar compo­nentes como filtros, siendo o no de la misma fuente de datos siempre y cuando los datos coincidan en los diversos conjuntos. Pueden ser utilizadas en la organiza­ciones para comprender rápidamente diferentes aspectos del negocio. También se puede utilizar para realizar proyecciones o tendencias, la cual Tableau nos ofrece de manera automática.

3.2. Aplicación de Data Discovery a datos de ins­tituciones del Estado

Generalmente las organizaciones no logran comprender en su totalidad los da­tos que generan. La consecuencia de no comprender esos datos puede resultar en la mala toma de decisión, lo cual podría ocasionar un gran impacto negativo a la organización. La información es considerada como uno de los recursos más im­portantes en una organización, y en base a esta información, se puede obtener conocimiento que podría ayudar a obtener mejores resultados.

En el presente trabajo son analizados datos de la ANDE y de la DGEEC, relacionando ambos conjuntos de datos, con el objetivo de obtener información de interés para la organización.

3.2.1. Datos de la ANDE y de la DGEEC

Se cuenta con datos de consumo de energía eléctrica, facturaciones, grupos de consumo (residencial, industrial, exportación, comercial, gubernamental y otros), por año (2000-2014), por departamento y distrito. Estos datos fueron solicitadosformalmente a la institución por medio de la Facultad de Ciencias y Tecnología de la Universidad Católica, a la cual tuvimos una respuesta favorable.

1. Dashboard de control / monitoramiento

En esta sección se presentan 4 productos construidos en esta Tesis, los cuales son Paneles de Control, en donde se relacionan conjuntos de datos de la ANDE y de la DGEEC. Una de las técnicas utilizadas para medir el crecimiento es la tasa de crecimiento, la cual se obtiene de la siguiente forma:

Es calculado el porcentaje de crecimiento ocurrido por cada ano (Ej.: Si al cerrar el año 2014, la cantidad de clientes fue de 1.000.000 y en el año 2015 aumento

1. esto quiere decir que, en el año 2015, la tasa de crecimiento de los clientes fue del 10 %, es decir, hubo un crecimiento positivo y la cantidad de clientes ha aumentado). Suponiendo que en el año 2016 la ANDE cierra con un total de
2. 000 de clientes, su crecimiento ser a 10 % menor al año anterior. La fórmula empleada (ver Formula abajo), donde n es el ano actual y n-1 el año anterior, PIB indica la cantidad de clientes que posee la ANDE.

formula

3.2.2.I. Dashboard - Clientes Facturados vs Crecimiento Poblacional

En el dashboard de la Figura 3.6 son utilizados datos históricos de la población, proveídos por la DGEEC, y datos de clientes, tales como el consumo y facturación, proveídos por la ANDE. Es importante notar en la figura, que inde­pendiente a que la tasa de crecimiento poblacional se mantenga casi constante, la tasa de crecimiento del consumo se eleva de forma pronunciada. Además, puede ser muy relevante la información de proyección del consumo para los próximos años, para elaborar una planificación en la ampliación de la capacidad de transmisión o distribución de energía.

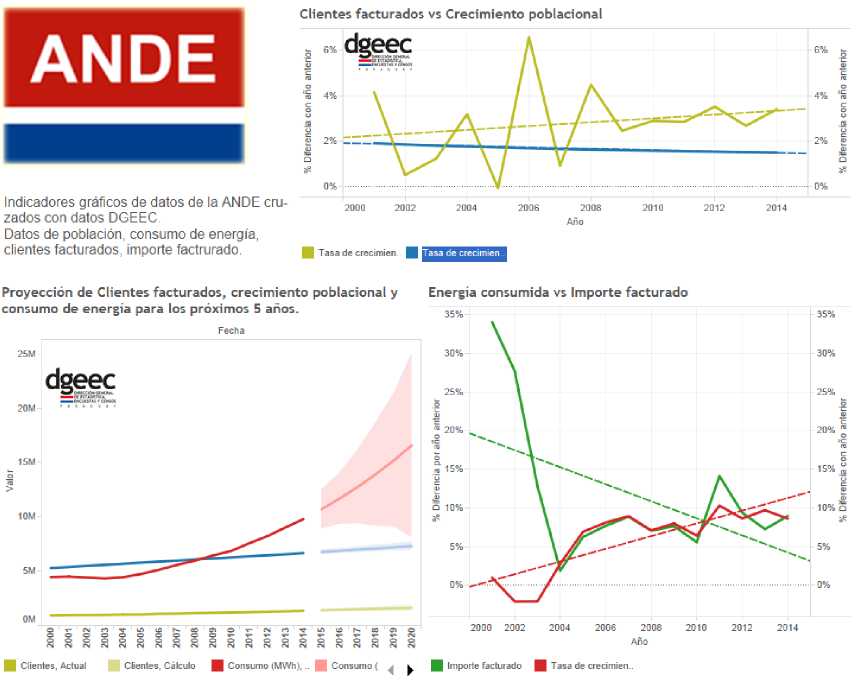


Figura 3.6: Clientes Facturados vs Crecimiento Poblacional. Elaboración propia.

En la Figura 3.7 se presentan dos tasas de crecimiento: de Clientes de la ANDE y de la Población del Paraguay. Es importante notar que el ritmo de crecimiento de la población disminuye con el tiempo, pero no así el ritmo de consumo de electri­cidad. No existe una relación directa entre estas dos tasas por el momento, lo cual puede deberse a factores tales como instalación de mayor cantidad de dispositivos eléctricos en residencias, o bien la instalación de más industrias las cuales tienen un alto nivel de consumo de electricidad. Es importante notar que la tendencia de la tasa de crecimiento del consumo de electricidad es positiva, lo cual podría ayudar a planificar una mayor inversión en la red de electricidad para aquellas zonas donde se registran tendencias mayores al consumo.

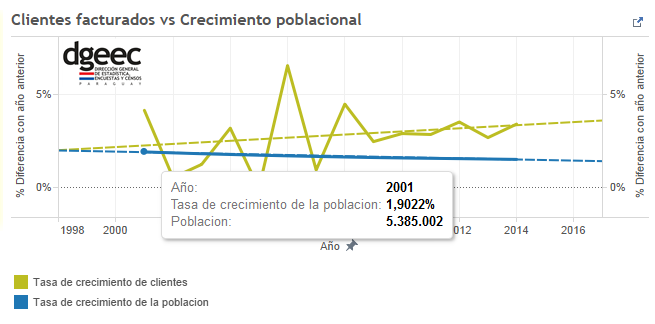


Figura 3.7: Clientes Facturados vs Crecimiento Poblacional. Elaboración propia.

En la figura 3.8 observamos que la cantidad de la población en el año 2002 cerró con un total de 5.484.610, la cual su crecimiento fue del 1,8497 % que equivale a 99608.

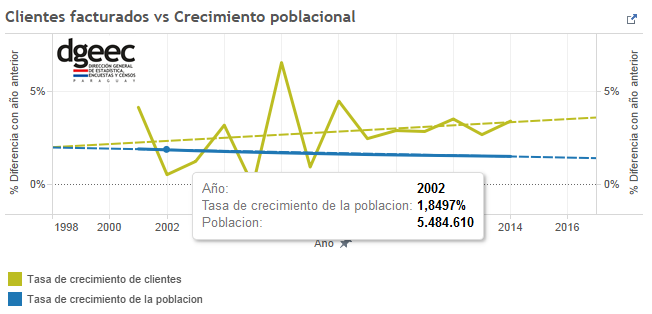


Figura 3.8: Clientes Facturados vs Crecimiento Poblacional. Elaboración propia.

En la figura 3.9 se muestran valores que representan el porcentaje del crecimien­to de los clientes de la ANDE. Como podemos ver, hay anos en que el aumento es muy evidente (2004,2006,2008) y hay años en que este es mínimo (2002,2005,2007).

Las líneas discontinuas representan las tendencias de ambos puntos. Por ejemplo, la cantidad de clientes en el año 2001 fue de 959.580, la cual aumento el 4.1 % respecto al año anterior.

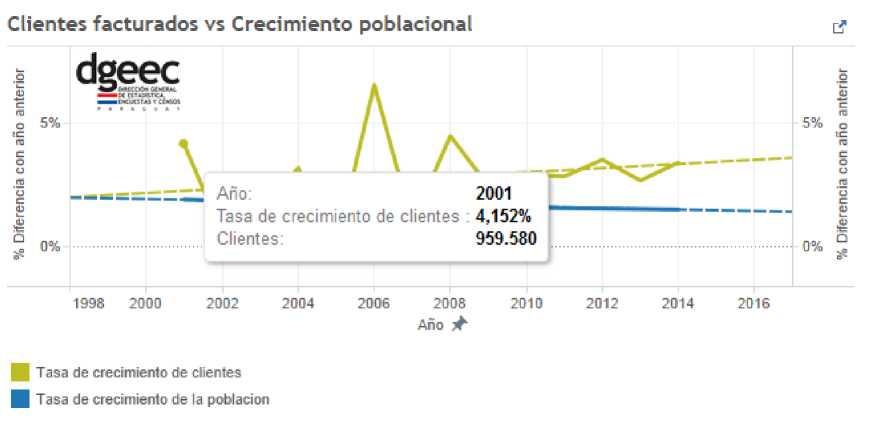


Figura 3.9: Clientes Facturados vs Crecimiento Poblacional. Elaboración propia.

En la figura 3.10, se muestra la cantidad de clientes correspondiente al año 2002, vemos que ascendió a 964.449 con un aumento de 4.869, que corresponde a un incremento del 0.5 % respecto al año 2001. Sin embargo, en la figura 3.11, en el año 2003 el incremento fue de 1.2 %, la cual representa a un aumento de más que el doble del año anterior llegando a aumentar 11830 clientes.

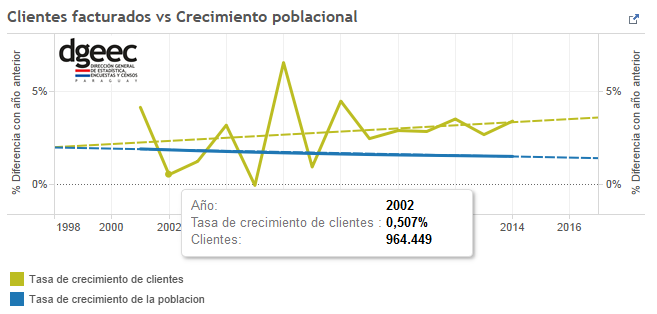


Figura 3.10: Clientes Facturados vs Crecimiento Poblacional. Elaboración propia.

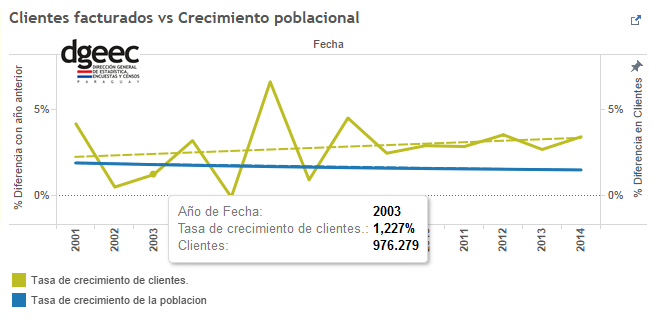


Figura 3.11: Clientes Facturados vs Crecimiento Poblacional. Elaboración propia.

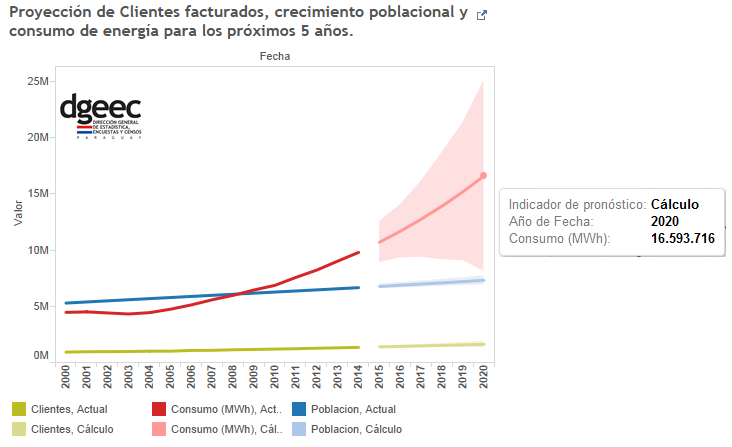
En la figura 3.12 se puede observar con más detalle la proyección (forecasting) del consumo, crecimiento poblacional y cantidad de clientes. Se puede notar que existe una mayor probabilidad que en promedio, el consumo de electricidad au­mente de forma considerable hasta el 2020 (línea roja). Además, este aumento es mayor en proporción al aumento de la población y a la cantidad de clientes.

Esto demuestra claramente que el consumo aumenta cada vez más, indepen­diente que se dé un aumento en la misma proporción de clientes o población. Esto

puede ocurrir a causa de varios factores, entre ellos, el aumento de dispositivos eléctricos en las residencias, debido al aumento en la capacidad adquisitiva de las personas. Tambien se puede dar a causa de un aumento en la cantidad de indus­trias.

Figura 3.12: Proyección de Clientes Facturados Crecimiento Poblacional Y Consumo De Energía para los Próximos 5 Años. Elaboración propia.

En el tercer y último gráfico de este panel, se muestra el porcentaje del creci­miento anual de los importes facturados y consumo de energía. Se puede observar que la facturación de la ANDE, en general es proporcional al consumo de energía eléctrica, excepto en el año 2011 (Figura 3.13), en la cual el importe facturado fue superior al aumento del consumo de energía. Sin embargo, en el año 2013 (Figura 3.14) se registra nuevamente un aumento en relación a años anteriores.



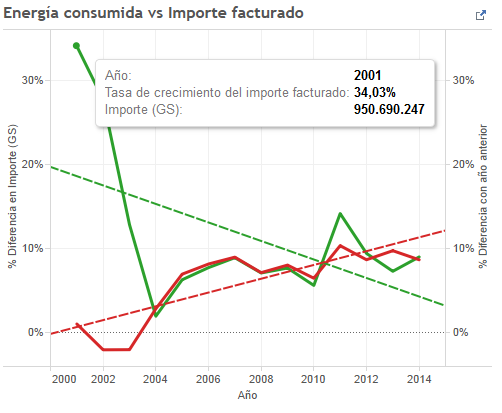


Figura 3.13: Energía Consumida vs Importe Facturado. Elaboración propia.

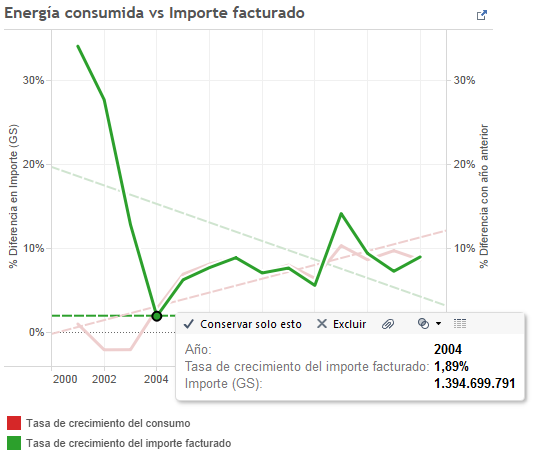


Figura 3.14: Energía Consumida vs Importe Facturado. Elaboración propia.

3.2.2.2. Panel estadístico de consumo de electricidad por sector

En la figura 3.15 se presenta un gráfico con el consumo de energía históri­ca, que abarca desde el año 1990 hasta 2014. Estos datos están clasificados por los siguientes criterios de la ANDE: Alumbrado Público, Comercial, Exportación, Gubernamental, Industrial y Residencial. En este gráfico se puede apreciar que el mayor consumo de energía se encuentra en el sector Residencial. Sin embargo, los valores de Exportación de energía fueron disminuyendo durante el tiempo. Esto tiene sentido debido a que ambos valores son inversamente proporcionales, esto es, cuando el consumo nacional se incrementa, se hace un mayor uso de energía en el país, por lo tanto, disminuye la energía disponible para exportación.



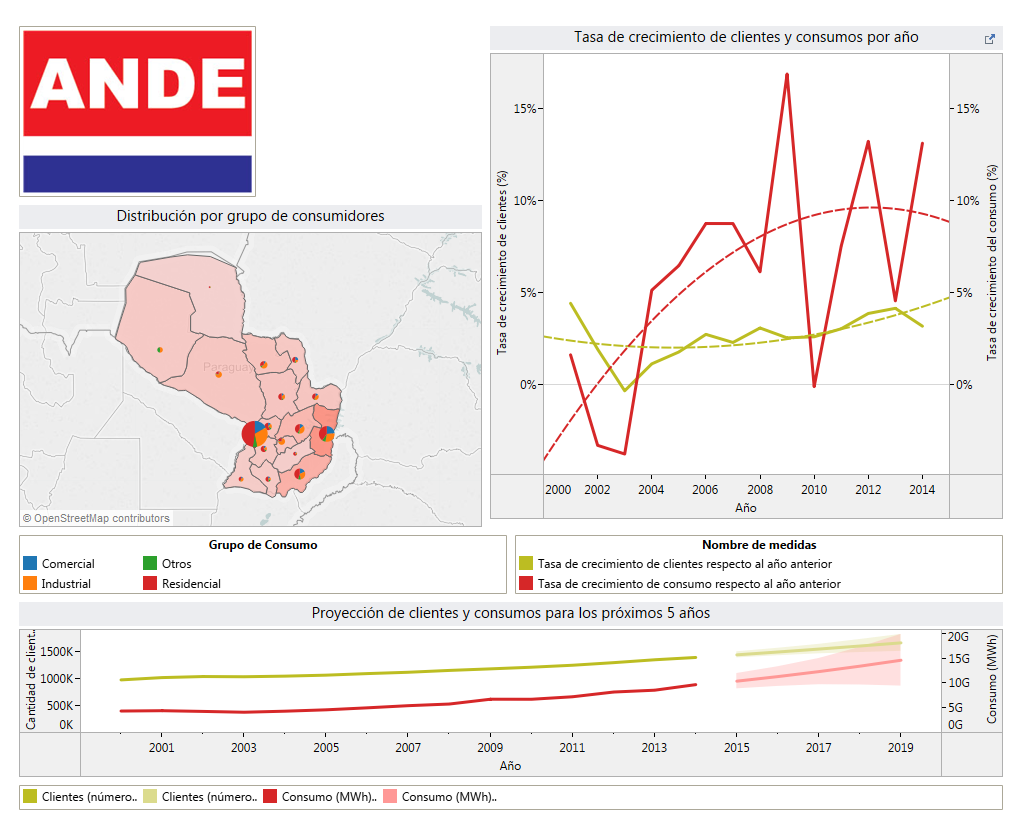
Figura 3.15: Estadística de consumo de electricidad por sector(1990-2014). Elaboración propia.

La Figura 3.16 presenta las informaciones de facturación también clasificados por sector, con el recurso de filtros por año. Es importante destacar un factor resaltante: aunque energía exportada disminuyo, el valor facturado por energía vendida al exterior aumento. Es probable que esto se haya debido al aumento de la tarifa de energía vendida, lo cual beneficia al país.

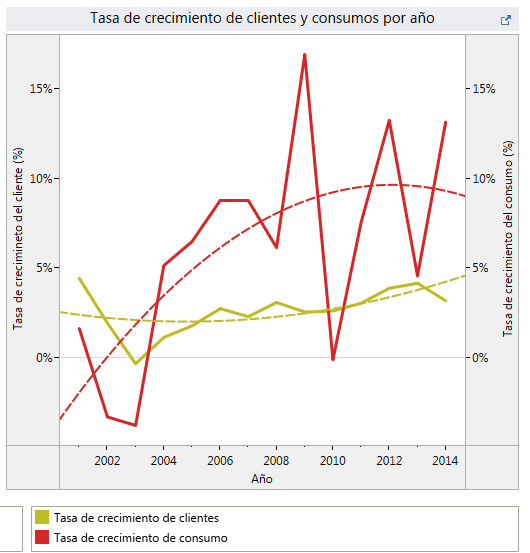
  
Figura 3.16: Importe facturado por año y sector(1990-2014). Elaboración propia.

1. Panel comparativo de Tasa de crecimiento y el Consumo de energía

En la figura 3.17 se presenta un panel comparativo entre la tasa de crecimiento de clientes y consumo de energía. También se cuenta con un mapa para filtrar por región del país.

Figura 3.17: Tasa de crecimiento vs Consumo de energía. Elaboración propia.

En el grafico situado a la derecha del mapa (Figura 3.18), se tiene el consumo y crecimiento de clientes. Para un mejor análisis fue necesario suavizar los datos calculando líneas de tendencia, debido a una inestabilidad de los datos de consumo. Así se puede apreciar, que existe una tendencia de crecimiento sostenido durante el tiempo de clientes. Sin embargo, la tendencia que el consumo crezca es mayor al de clientes.

Figura 3.18: Tasa de crecimiento vs Consumo de energía, filtrado por el departa­mento Alto Paraná. Elaboración propia.

Debajo se presenta un cuadro con la línea de cada valor (Figura 3.19), de creci­miento y consumo. Con el uso de un recurso disponible que cuenta la herramienta escogida Tableau, es posible realizar análisis predictivo (forecasting) del crecimien­to y consumo. Tableau utiliza un algoritmo llamado Suavizado Exponencial, muy conocido en el área de Matemáticas Estadísticas. En este gráfico se puede notar que existe una mayor probabilidad que en los próximos años aumente considera­blemente el consumo, superando su media. Sin embargo, se nota que el ritmo de crecimiento de clientes es sostenible, y no tiene una alta probabilidad de sufrir un aumento abrupto.



Figura 3.19: Proyección de clientes y consumos para los próximos 5 años. Elaboración propia.

1. Panel de Tasa de crecimiento poblacional y Consumo de ener­gía anual

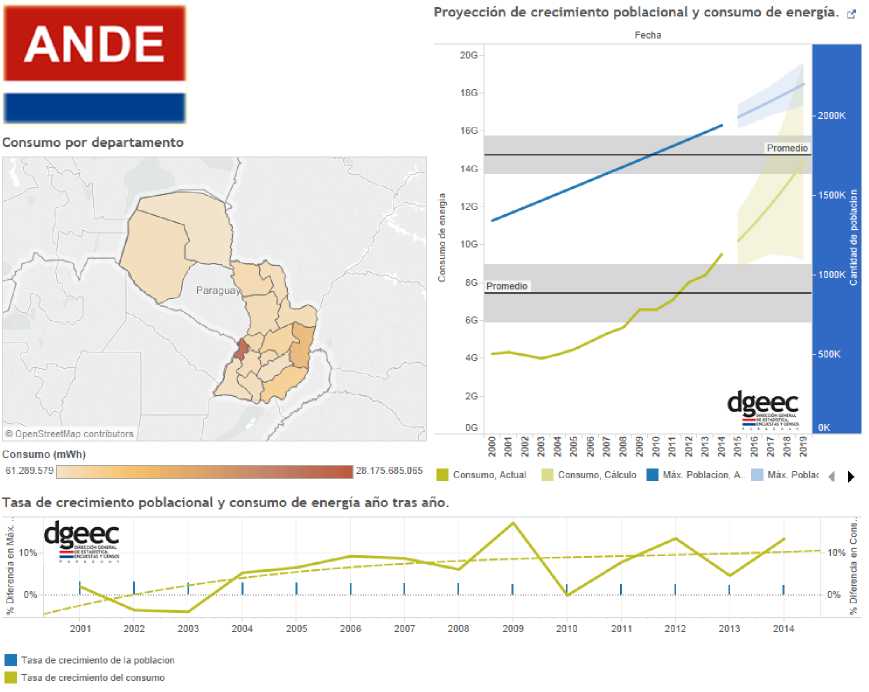


Figura 3.20: Panel de Tasa de crecimiento poblacional y Consumo de energía anual. Elaboración propia.

En la figura 3.21 se presenta el panel comparativo de datos de crecimiento poblacional de la DGEEC y de consumo de la ANDE. En el mapa, cuando el color es más oscuro el consumo de energía es mayor, y si el color es más claro, el consumo es menor. Se puede notar que los departamentos Central y Alto Paraná son los que tiene mayor consumo de energía. Este tipo de gráfico es muy útil cuando se desea analizar información de forma general y georreferenciada. Al ubicar el mouse sobre cualquier departamento, se presenta una ventana emergente indicando el valor de consumo del departamento seleccionado. Al seleccionar un departamento del mapa, los demás gráficos también se actualizan filtrando el departamento seleccionando.

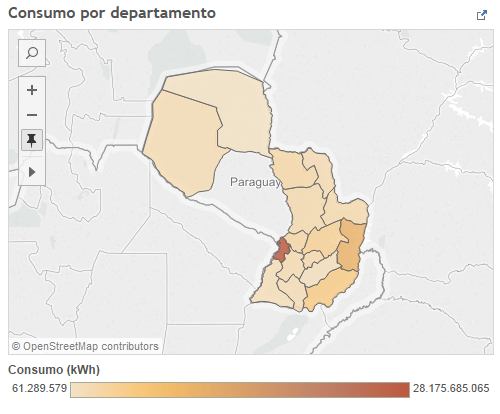


Figura 3.21: Consumo por departamento. Elaboración propia.

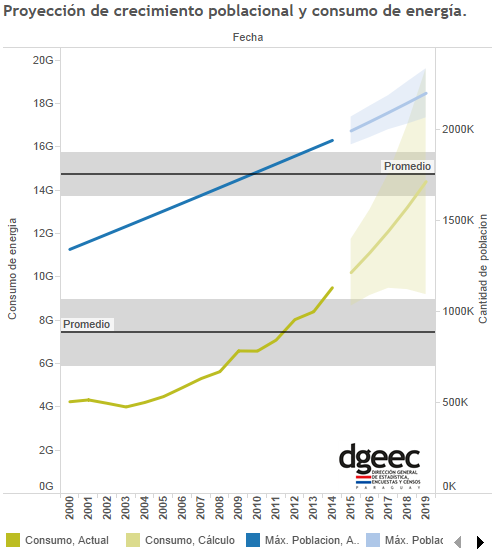
Los gráficos situados a la derecha y debajo del mapa presentan la tasa de crecimiento poblacional y de consumo de energía. Es importante notar que la tasa de crecimiento de la población es casi constante, esto es, no tiene una gran variación en el tiempo, ni una tendencia a alejarse de la media. Sin embargo, la tasa de crecimiento de consumo tiene una línea de tendencia a crecer durante eltiempo. Así puede apreciarse que, el consumo de energía no tiene una relación de proporcionalidad con respecto al crecimiento de la población.

Figura 3.22: Proyección de crecimiento poblacional y consumo de energía. Elaboración propia.

En este gráfico, se muestra la misma información que el gráfico anterior, pero con diferente perspectiva, en este caso se calcula el porcentaje de crecimiento anual tanto de la población, así como del consumo.

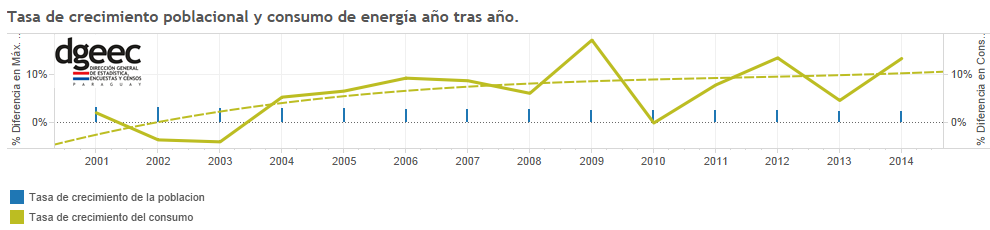


Figura 3.23: Tasa de crecimiento poblacional. Elaboración propia.

1. Marco Metodológico
2. Alcance

Aplicaremos las técnicas de Data Discovery a los datos de dos instituciones del estado, específicamente la ANDE y DGEEC, donde demostraremos que con datos de calidad podríamos detectar oportunidades que nos faciliten la toma de decisiones en la institución. Utilizaremos conjuntos de datos de las instituciones mencionadas más arriba para este fin.

1. Enfoque

El enfoque que utilizamos es el cuantitativo, que, por lo común, utiliza la recolección y el análisis de datos para contestar preguntas de investigación y probar hipótesis establecidas previamente, y confía en la medición numérica, el conteo, y en el uso de la estadística para intentar establecer con exactitud patrones en una población. (por ejemplo, un censo es un enfoque cuantitativo del estudio demográ­fico de la población de un país).

1. Técnica e Instrumentos de recolección de datos

La técnica aplicada en este trabajo en la recolección de datos fue la investigación de documentos científicos procedentes de publicaciones de empresas pioneras en Data Discovery y de expertos en el área.

1. Conclusiones y Trabajos futuros

Este trabajo fue dividido en las siguientes partes:

* Capítulo 1: en este capítulo fue presentada una introducción a los objetivos de este trabajo.
* Capítulo 2: Marco Teórico: esta sección desarrolla el estado del arte en el área de BI y principalmente el de Data Discovery, técnica aplicada en este trabajo.
* Capítulo 3: En esta sección se presenta la evaluación técnica realizada para la selección de la herramienta Tableau. Además, la aplicación de técnicas de Data Discovery a los datos de la ANDE y la DGEEC. También se presentan los productos construidos en este trabajo, para el análisis de las informaciones, y las proyecciones realizadas para los próximos anos.
* Capítulo 4: En esta sección se presenta el marco metodológico utilizado para el desarrollo de esta tesis, incluyendo el alcance, el enfoque y la técnica e instrumentos de recolección de datos.

Para el desarrollo de este trabajo se obtuvieron datos de la ANDE y la DGEEC. Con estos datos fue posible aplicar técnicas de Data Discovery realizar un análisis de las informaciones, cruzarlos, georreferenciarlos, encontrar líneas de tendencias y pronósticos de crecimiento a futuro tanto del consumo de energía, de clientes de la ANDE y de la población del país. En cada caso se presenta un análisis que demuestra con gráficos intuitivos que en ciertas ocasiones no existe una relación proporcional entra algunas dimensiones. Sin embargo, teniendo en cuenta los re­sultados obtenidos se pueden observar las siguientes cuestiones:

1. La tasa de crecimiento del consumo es mayor a la tasa de aumento de clientes.
2. Debido a este aumento en el consumo de energía del país, disminuyo la cantidad de energía exportada.
3. Aunque se tuvo una disminución en la energía exportada, se obtuvo un cre­cimiento en el valor facturado. Esto demuestra una mejoría en el precio de venta de la energía al Brasil (Itaipú) o Argentina (Yacyretá).
4. Según el pronóstico de crecimiento para los próximos años, el consumo de energía tendría un crecimiento mayor al de la cantidad de clientes, y la población del país.
5. La tasa de crecimiento de la población se mantiene constante durante el tiempo, esto es, la población crece a una tasa sostenida. Sin embargo, la tasa de aumento en el consumo de energía es considerablemente mayor, y

demuestra un aumento abrupto para los próximos años, independiente a la tasa de aumento de la población y de nuevos clientes de la ANDE.

Teniendo en cuenta estas cuestiones, este trabajo puede servir de herramienta para la Planificación en la Inversión en la capacidad de transmisión y distribución de electricidad dentro del territorio del país, para los próximos años, además de ayudar con indicadores para montar una estrategia de exportación para replantear las condiciones actuales de venta de la energía al exterior.

49

Referencias

Adelman, S., Moss, L., y Barbusinski, L. (2002). I found several definitions of bi. DM Review, 5700-1.

Davenport, T. (1993). Process innovation: reengineering work through information technology. Harvard Business Press.

Denison, D. R. (1997). Toward a process-based theory of organizational design: Can organizations be designed around value chains and networks? Advances in Strategic Management, If, 1-44.

Eckerson, W. (2009). Who ensures clean, consistent data. The Datawarehouse Institute.

Gangadharan, G. R., y Swami, S. N. (2004). Business intelligence systems: design and implementation strategies. En Information technology interfaces, 2004. 26th international conference on (pp. 139-144).

Goebel, M., y Gruenwald, L. (1999). A survey of data mining and knowledge discovery software tools. ACM SIGKDD explorations newsletter, 1 (1), 20­33.

Hancock, J. C., y Toren, R. (2006). Practical business intelligence with sql server 2005. Pearson Education.

Herschel, G., Linden, A., y Kart, L. (2015). Magic quadrant for advanced analytics platforms. Gartner Report G, 270612.

Josh Parenteau, C. H., Rita L. Sallam. (2015). The rise of data discovery has set the stage for a major strategic shift in the bi and analytics platform market. Gartner Research. Gartner Inc, 1.

Kumari, N. (2013). Business intelligence in a nutshell. International Journal of Innovative Research in Computer and Communication Engineering, 1 (4), 969-975.

Malhotra, Y. (2001). From information management to knowledge management. beyond the’hi-tech hidebound’systems. Knowledge management and business model innovation, 115-134.

Marakas, G. M. (2003). Modern data warehousing, mining, and visualization: core concepts. Prentice Hall.

Nguyen, T. M., Schiefer, J., y Tjoa, A. M. (2005). Sense & response service architecture (saresa): an approach towards a real-time business intelligence solution and its use for a fraud detection application. En Proceedings of the 8th acm international workshop on data warehousing and olap (pp. 77-86).

Ranjan, J. (2009). Business intelligence: concepts, components, techniques and benefits. Journal of Theoretical and Applied Information Technology, 9 (1), 60-70.

Rud, O. P. (2009). Business intelligence success factors: tools for aligning your business in the global economy (Vol. 18). John Wiley & Sons.

Sallam, H. S. O. T. H., Parenteau. (2015). *Critical capabilities for business inte­lligence and analytics platforms.* Gartner Research. Gartner Inc, 1.

Seufert, A., y Schiefer, J. (2005). *Enhanced business intelligence-supporting bu­siness processes with real-time business analytics. En Database and expert systems applications, 2005. proceedings. sixteenth international workshop on* (pp. 919-925).

Tapadinhas, J. (2014). *How to Architect the BI and Analytics Platform*. Gartner Research. Gartner Inc, 1.Material no publicado.

Peck, G. (2013). *Tableau 8: The Official Guide.* McGraw-Hill Osborne Media.