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**

II

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

IV

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

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

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

Tutor: Ing. Ricardo Luis Brunelli Montero

vi

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 Catálica. 110 p.

Tutor: Ing. Ricardo Luis Brunelli Montero.

Defensa de Proyecto de Fin de Carrera.

Palabras clave: Data Discovery, Business Intelligence.

vil

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

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

Mesa Examinadora

Prof. Nelida 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 comprension 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, empeno, 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 facil 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, empeno, perseverancia, y  
coraje para conseguir nuestros objetivos.

Ariel Hernán Landaida Duarte

XII

RESUMEN

Hasta hace poco tiempo, la mayoría de las organizaciones proveían datos es­tructurados, limpios, e integrados, resumidos a niveles convenientes para plata­formas convencionales. Data Warehouse 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 heterogeneos y hasta inclusive no estructurados, y esa situación generó la necesidad de crear otro tipo de herramienta que ayude al tomador de decision en la busqueda de patrones y relaciones. Este nuevo enfoque, denominado Data Discovery, no podía ser igual a las tecnicas ya tradicionales, ademós debía tener características como innovacion 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 area de BI y específicamente Data Discovery. Se aplican estas tecnicas a datos de dos instituciones del estado, demostrando los beneficios de aplicar este tipo de tecnica.

Palabras clave: Data Discovery, Business Intelligence, Data Warehouse.

xiv

ABSTRACT

Until recent time, most organizations provide structured, clean, and integrated data, summarized in desirable levels for conventional platforms. Data Warehouse 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, Data Warehouse.

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)

Proyección de crecimiento poblacional y consumo ee energía 44

3.22.

**lista de tablas**

2 1. Técnicas actuales de BI

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 Analisis 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

xx

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 tecnicas 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 institution. Utiliza­remos conjuntos de datos de las instituciones mencionadas mías arriba para este fin. Este trabajo esta estructurado de la siguiente forma: El capítulo I presenta el planteamiento del problema, los objetivos generales y específicos, concluyendo con la justification. El capítulo II presenta el marco teorico. 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 magico de Gartner, el cual apoyo a la selection de la herramienta mencionada. Ademas se presentan ejemplos de analisis de da­tos y la información obtenida, como resultado de la combination del uso de las herramientas y datos.

1. Planteamiento del problema

Actualmente la mayoría de las instituciones publicas e incluso las del sector pri­vado, tienen un bajo nivel de inversion en tecnología. Generalmente en empresas que exigen toma de decisiíon, con frecuencia optan por decisiones de negocios no íoptimas debido a que no poseen la suficiente experiencia o suficiente datos procesa­dos del negocio para llegar al correcto analisis, o pueden estar usando herramientas incorrectas. Teniendo ínicamente la experiencia como herramienta, puede ser su­ficiente sílo en organ

izaciones pequeñas, donde se adquiere conocimiento sin necesidad de alguna herramienta de analisis de datos, por ejemplo, sabriamos cuales son los productos mas vendidos o mas 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 infor­mation. Estas informaciones son obtenidas a traves de un analisis profundo de datos, por consiguiente estos datos tambien deben ser de buena calidad. Si una organization 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 evolucion y estabilidad a lo largo del tiempo.

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

1. Objetivos
2. Objetivo General

En este trabajo se aplican las tecnicas 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 interes para una organizaciáon.

1. Objetivos Específicos

* Estudio del estado del arte de BI y Data discovery.
* Selection de la herramienta adecuada para este trabajo.
* Analisis y cruzamiento de datos para el descubrimiento de situaciones de interes.
* Elaboration de dashboards que reflejen el análisis de los datos con la herra­mienta seleccionada.
* Elaboracián de gráficos de tendencias, proyecciones de consumo y clientes de la instituciáon.

1. Justificación

Falta de un ambiente analítico corporativo, que proporcione informaciones e indicadores necesarios para la toma de decision. Falta de agilidad para elaborar informes con indicadores de tendencias de consumo de energía, que apoye a la planificación territorial de expansion de la red de transmisión electrica. Falta de indicadores de consumo geográfico de electricidad, relacionados con indicadores po- blacionales, para apoyo en la planificacion de inversion en el aumento de capacidad de transformadores por region.

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 information correcta pa­ra la persona correcta, en el tiempo correcto, a traves del canal correcto (Kumari, 2013).

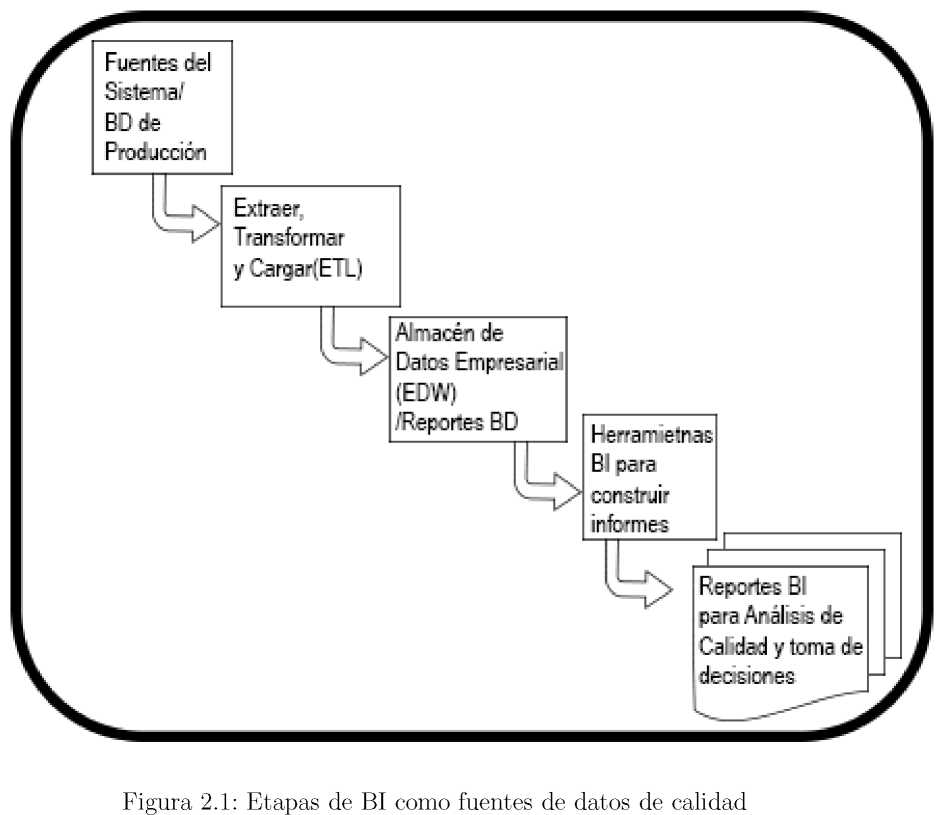
Esto produce grandes cantidades de information que pueden llevar al desarrollo de nuevas oportunidades para la organization. La identificación e implementation de estas oportunidades en un estrategia efectiva, puede proporcionar una ventaja competitiva de mercado y la deseada estabilidad a largo plazo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

Las tecnologías de BI ofrecen vistas históricas, actuales y predictivas de las ope­raciones de una empresa. Algunas de las funcionalidades mós comunes de BI son reportes, procesamiento analítico en línea (en adelante OLAP, Analytical Proces­sing), minería de datos, minería de procesos, procesamiento de eventos complejos, gestion del rendimiento empresarial, benchmarking, minería de texto, anólisis pre- dictivos y analisis prescriptivos. BI tiene como objetivo apoyar una mejor toma de decisiones, por tanto, un sistema de BI puede tambien ser llamado sistema de apoyo a las decisiones o DSS (del ingles, decision support system)(Rud, 2009).

Las herramientas de BI son disenadas para obtener, analizar y reportar datos. Estas herramientas generalmente leen datos que han sido previamente almace­nados, frecuentemente, aunque no necesariamente en un almacen 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, metodos, y tecnologías para convertir datos separados o aislados de una organization en information util 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 organizaciín. 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 organization (Kumari, 2013).



Una cuestion fundamental es el hecho que una organization corre sobre datos; y actúa como insumo para el motor de la industria corporativa. Una organization 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 direction de una empresa y el area de TI (tecnología de la information) 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 navegacion de dimensiones tales como el tiempo o jerarquías. OLAP, provee vistas resumidas multidimensionales de datos del negocio de una organization, y es usado para reportes, analisis, modelado y planificacion para la optimización de una organizacion.(Malhotra, 2001)

Las tecnicas y herramientas OLAP pueden ser usados para trabajar con da- tawarehouse o con data marts (un subconjunto de datos de un órea 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 gestion 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; almacen de documentos y gestion de documentos; gestion del conocimiento; mapeamiento, visualización de information y paneles (dashboards); sistemas de information de gestion, sistemas de información geográficas; anólisis de tendencias; Software como servicio (SaaS) y otros.(Malhotra, 2001)

1. Análisis Avanzado Analisis Avanzado: conocido como minería de datos y analisis predictivos, toma las ventajas de las tecnicas de analisis estadís­ticos para predecir o proveer medidas de certeza sobre ciertos hechos. La gestion 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 móetricas y patrones de aprendizaje y crecimiento en las organizaciones.(Gangadharan y Swami, 2004)
2. BI en tiempo real: permite la distribution de metricas en tiempo

real a traves de emails, sistemas de mensajería instanránea y/o pantallas interac­tivas.

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

Tambien puede contener datos operacionales, los cuales pueden ser definidos como un conjunto actualizable de datos integrados disponibles para toda una or- ganizacion, para la toma de decisiones rácticas 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 organization. (Gangadharan y Swami, 2004)

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

Cada departamento tiene su propia interpretación de cómo debe verse un data mart y el data mart de cada departamento es particular y atiende las necesida­des específicas del area. Similar al datawarehouse, los data marts contienen datos transaccionales que ayudan a expertos en negocios a crear una estrategia basada en el analisis de las tendencias y experiencias pasadas. La principal diferencia es que la creación de los data marts se basa en una necesidad específica, predefinida para un grupo determinado. Un data mart 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 decision, estas estón desacopladas y extraen informaciones de transacciones de negocio. Las habilidades de BI incluyen, apoyo a la decision, procesamiento analítico en línea, analisis estadísticos, analisis 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 investigation de mercados o desde internet, datos no estructurados de redes sociales, o information desde un datawarehouse existente. Las fuentes de datos pueden ser bases de datos relaciona­les o cualquier otra estructura de datos de sistemas transaccionales. Ellos tambien pueden residir en muchas plataformas diferentes, tales como tablas, hojas de calcu­lo, o information no estructurada, tales como archivos de texto plano o imagenes 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 co- municacion 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 information es frecuentemente considerada como el segundo recurso mas

importante que una compañía tiene (lo más valorable de una compañía son las personas). Cuando una compama puede tomar decisiones basadas en información oportuna y precisa, puede ayudar a mejorar su rendimiento en su segmento de mercado. BI tambien agiliza la toma de decisiones, ayuda a actuar rápida y correc­tamente con la informacioán adecuada antes de otras empresas de la competencia. Tambien 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 facilmente convertir sus cono­cimientos de negocio en inteligencia analítica para resolver muchas cuestiones de negocio, tales como incrementar la tasa de respuestas desde correos elec- tronicos, telefonos, y mejorar las campañas de ventas desde internet.
* Con BI, las empresas pueden identificar sus clientes mas 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.
* Detection rápidamente de problemas reportados de productos para minimi­zar el impacto de las deficiencias en sus disenos.
* Descubrir lavado de dinero de actividades delictivas.
* Analizar la rentabilidad potencial del cliente, y reducir el riesgo a traves de una puntuacion mas precisa de crédito financiero de los mismos.
* Determinar cuales son las combinaciones de productos y servicios que los clientes son mas propensos a comprar y cuando.
* Analizar los ensayos clínicos de farmacos experimentales.
* Establecer tarifas mas rentables para las primas de seguros.
* Reducir el tiempo fuera de un equipamiento mediante la aplicacián de man­tenimiento predictivo.
* Determinar con el analisis de desercián y rotation 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 farmacos moleculares prometedores.

1. Tecnología de BI

La inteligencia empresarial provee datos organizacionales de tal manera que los filtros de conocimientos organizacionales puedan facilmente asociarse con estos datos y volverlos en informacián para la organization. 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 data warehouse porque es un conjunto de information logica recolectado desde varias bases de datos operacio- nales con el objetivo de crear inteligencia de negocios (Ranjan, 2009).

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

* Seguridad y acceso de usuarios al data warehouse.
* Volumen de datos (capacidad).
* Cuanto tiempo sera almacenado el dato (retention 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) - Logica Asociativa de Consultas.
* Metricas y mediciones del rendimiento del negocio.
* Planeamiento Empresarial.
* Data mining (DM), Data Farming, y Data warehouses.
* Sistemas de apoyo a la decisión (DSS) y prediction.
* Datawarehouse de documentos y gestion documental.
* Sistema de Gestion Empresarial.
* Finanzas y presupuestos.
* Recursos humanos.
* Gestión del conocimiento.
* Mapeamiento, visualización de la information, y paneles de control (dash­boards).
* Sistemas de gestion de informaciones.
* Sistemas de information geográfica (GIS).
* OLAP (Online Analytical Processing) y analisis multidimensional; a veces simplemente llamado “Analytics“ (basado tambien en “hipercubo“ o “cubo“).
* BI en tiempo real.
* Analisis de datos estadísticos y tecnicos.
* Gestion de la línea de producción, Gestión de demandas.
* Gestion de la cadena de Suministro/Gestión de la cadena de demanda.
* Analisis 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 mas organizaciones han comenzado a disponibilizar más datos con mayor velocidad. En el pasado, los datos solo estaban disponibles despues 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 estan disponibles a cada 24 horas, requiriendo la automatization y el uso de sistemas de TI.

1. Breve discución

La experiencia actual de cualquier nueva forma de organizacián 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 termino que engloba un amplio rango de software de analisis y soluciones para recolectar, consolidar, analizar y proveer acceso a la informacioán 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 organization, trayendo information en tiempo real para centralizar repositorios y apoyar el analisis, 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 aplicacián, asá como practicas de analisis (Gangadharan y Swami, 2004). BI es técnicamente mas amplio, lo que potencialmente engloba la gestiáon del conocimiento, la planificacioán 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 reaction 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 traves de procesos de negocios corporativos, y brindar recomendaciones automáaticamente para optimizar las ope-

raciones 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 informacion 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 cuantas mantas fueron vendidas en un mes, las empresas desean entender cuóantas fueron vendidas en un país determinado durante un evento meteorológico. BI proporcio­na una vision integrada unificada de las actividades empresariales. Las empresas han construido sistemas de BI que apoyan anóalisis 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 estón 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 analisis de BI, herramientas y modelos de datos que permiten a una organization aprovechar su informacion. Las herramientas BI facilitan el acceso a un gran volu­men de datos corporativos, y convertir esos datos en informacion util y procesable que sea consistente a traves de la version coherente de la verdad. Las empresas auón sienten que BI tiene complejidades relacionadas con la tecnologóa y que pue­de usarse solamente por especialistas con conocimientos tecnicos, ademas que los costos de implantation son altos. Las empresas requieren estos analisis 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 aplicacion de apoyo a la decision (incluyendo data warehouses y data marts) son BI, si y solo si fueron desarrollados bajo la proteccioón y la metodologóa de una iniciativa estratóegica de toda la Organization (Gangadharan y Swami, 2004). Los sistemas tradicionales de BI consisten en una base de datos en el back-end, una interfaz de usuario en el front-end, software que procesa la informacióon para producir la propia inteligen­cia de negocios, y un sistema de informes. Las capacidades de BI incluyen apoyo a la decision, procesamiento analítico en linea, analisis estadísticos, prediction y minería de datos. Diferentes sectores como fabricantes, comercios electrónicos, em­presas de telecomunicaciones, aerolíneas, minoristas, sistemas de salud, servicios financieros, bioinformítica y hoteles utilizan BI para apoyo a clientes, investiga­tion de mercado, segmentation, rentabilidad del producto, analisis y distribution

de stock, análisis estadístico, informes multidimensional, detección fraudes, en­tre otros. BI y minería de datos es un area que esta fuertemente influenciado por tecnicas estadísticas tradicionales, y la mayoría de los metodos de minería de datos revela una fuerte base de metodos estadísticos y de analisis de datos. Algunas de las tecnicas tradicionales de minería de datos incluyen clasificacion, agrupación, analisis de valores atípicos, patrones secuenciales, analisis de series temporales, la prediction, la regresion, analisis de enlaces (asociaciones), y metodos multidimen­sional incluyendo el procesamiento analítico en línea Procesamiento nalitico en linea (OLAP). Estos pueden clasificarse en una serie de tecnicas 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. |
| Caracterizacion y minería de datos des­criptivo | Distribution, dispersion y exception de datos |
| Asociaciín, correlaciín, anílisis de la causalidad (Analisis Link) | Identificar relation entre atributos |
| Clasificacion | Determinar a que clase pertenece un elemento de datos |
| La agrupacion y analisis de valores atí­picos | Particiín de un conjunto en clases, con lo cual elementos con características si­milares se agrupan |
| Anílisis de patrones temporal y se- cuencial | Tendencia y desviacion, patrones se­cuenciales, frecuencia |
| OLAP(Procesamiento Analitico en Li­nea) | Herramientas OLAP permiten a los usuarios analizar distintas dimensiones de datos multidimensionales. Por ejem­plo, proporciona series temporales y puntos de vista de aníalisis de tenden­cias. |
| Modelo de visualization | Hacer fícil la descubierta de cono­cimiento usando charts, plots, histo­grams y otros medios visuales |
| Analisis Exploratorio de Datos(EDA) | Explorar un conjunto de datos sin una fuerte dependencia en hipotesis o mo­delos; el objetivo es identificar patrones de una manera exploratoria |

Cuadro 2.1: Tecnicas actuales de BI

En el siguiente capítulo se presenta una introducción a los métodos y técnicas de análisis exploratorio, y en especial la tecnica 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 multiples 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 semantica y metadatos, la cual es generalmente accesible solo por herramientas del propio fabricante. La situation actual de carácter propietario de la capa semantica 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 tecnicos y la complejidad de las estructuras de datos subyacentes. Sin embargo, con la adoption y rápido crecimiento de las herramientas de data discovery como Qlik, Tableau, y Tibco Spotfire, los usuarios buscan cada vez mas acceso a datos confiables en capas semáanticas cerradas, y los proveedores de BI se enfrentan a un reto difícil para seguir siendo relevantes en un mercado que está en transition. La respuesta de los proveedores tradicionales y la inversion significativa hasta la fecha, ha sido la utilización de sus capas semanticas existentes para promover la gestion 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 exito 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 semantica que sirve como la base de la mayoría de las plataformas BI tradicionales ha sido ampliamente adoptada por muchas organizaciones a traves 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, operation y mejora de estas plataformas construidas alrededor de una definida y centralizada capa semaíntica propietaria.

Este enfoque ha funcionado bien cuando el objetivo fuera una unica 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 presentation 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 semantica 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 mas abierto para sus capas semanticas propietarias, y esto puede llevar a un cambio mayor de mer­cado con el tiempo. La oportunidad probablemente sera dictada por el exito 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 inedito en BI y en mercados analiticos. Ejemplos incluyen Oracle OBI EE, Microsoft Power, SAP BusinessObjects y, mós recientemente, conectividad nativa de Tableau para modelo de datos Birst a traves de su capa semantica propietaria.

Oracle BI Enterprise Edition(OBI EE) fue uno de los primeros productos de plataformas BI para tener una capa semantica 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 conexióon ODBC que puede ser consumida por herramientas e interfaces de terceros. Inicial­mente, esto es cóomo Oracle proporciona conectividad a su herramienta interna de informes de produccion desarrollada, Oracle BI Publisher. Mós alla 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 mas 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 habóa permitido el acceso a su estructura de datos propietaria o capa semantica, hizo accesible sóolo a de informes, dashboard y capacidades data discovery auto- contenidas dentro de la cartera Birst. A traves de la alianza y desarrollo conjunto, se anadio una conexion con Tableau que permite la conexion directa con el medio ambiente Birst. Esto proporciona una mayor flexibilidad y una gama mas 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 semantica 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 semantica como una forma de promover a proveedores independientes que rigen capacidades de data discovery. Esto es, sin embargo, una oportunidad potencial de modernization 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 traves de servicios web.

Mientras que los clientes prefieren una única solution de sus proveedores titu­lares, ya sea de los proveedores de plataforma tradicional de BI o de proveedor de data discovery, los clientes actualmente encontraráan máas oportunidades para acceder a la capa semantica 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 Analisis de Negocio de Gartner, el cual se describe a continuaciáon.

Hay una serie de defectos relacionados en la mayoría de las organizaciones con relaciáon a BI y plataformas analiticas, asá como una percepciáon erroánea 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 solution estrechamente inte­grada con pocos componentes, preferentemente de un ánico proveedor, para entregar una sola version de la verdad para la organization.
* La information puede ser confiable sálo si está almacenada en un data wa­rehouse corporativo y entregada a los consumidores de informacioán usando artefactos de BI, tales como informes y dashboards.
* Las informaciones creadas o manipuladas por los usuarios de negocios inevi­tablemente producirán discrepancias a traves de diferentes analisis, lo que lleva a decisiones equivocadas, generando caos en la organizacián a traves del tiempo.
* La responsabilidad del departamento TI para la gestion de information ter­mina en la capa semáantica de BI y en los contenidos orientados a TI.
* Procesos analáticos orientados al negocio esrán fuera del alcance y no soporta­dos por TI. Hay varios problemas orientados a la informaciáon documentados en el mundo BI y análisis, que obligan a los láderes de BI a seguir estas creencias y desplegar un entorno de BI centralizado y monolático, que termi­na siendo impuesta a los usuarios, independientemente de su adecuaciáon a las necesidades.
* Los proveedores licenciados de BI, a favor de sus propias plataformas, apoyan

este enfoque.

Según Gartner, una plataforma analítica necesitan 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 proposito es ayudar a los usuarios a alcanzar sus objetivos de negocio a traves 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 analisis resultante tambien exigiró cambios en los procesos de gestion de la information, como atribuir nuevas responsabilidades a diferentes personas en la organización.

A continuation el framework de analisis de negocios. La figura 2.2 presenta el framework de analisis de negocios de Gartner.

**Desempeño**

**Modelos de Negocio, Estrategia de Negocio, y Métricas Empresariales**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Consumir |  | Procesos  de |  | Capacidades de Decisión Colaboración, Toma de Decision, |
|  |  |  | Decisión |  | Automatización de la Decisión |
|  |  |  |  | Plataformas | Inteligente, Aplicaciones |
| Persona | Producir | Procesos | Procesos  Análiticos | Capacidades Analíticas Descriptivo, Diagnóstico,  Predictivo, Prescriptivo |
|  |  |  |  |  |
|  | Habilitar |  | Procesos de la Gestión de Información |  | Capacidades de Información Describir, Organizar, Integrar, Compartir, Gestionar. Implementar |

Gestión de Programa

***O***

**£**

<D

**m**

**>.**

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| (m  Información | □ | 0 | te | /A.  m | & 8\* | 6 et |
| Social | Documentos | Transacciones | IT/OT | Imagen | Audio Texto Video | Móvil Búsqueda |

Figura 2.2: Framework Analitico de Negocios Gartner

ponentes de la plataforma que apoyan la transformación de la información en un mejor rendimiento de la organization. El uso de esta herramienta esta hecha por la lectura desde arriba hacia abajo, comenzando con los resultados del negocio y luego descifrando las composiciones analíticas de apoyo y la information necesaria para alcanzarlos. De acuerdo a las necesidades de los usuarios, la plataforma debe ser redisenada con un amplio conjunto de capacidades tecnicas (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 tambien muy ótil 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 organization muy probablemente tambien necesitara re-organizar y capacitar a los proveedores y usuarios de BI y anólisis. 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 enfasis en las plataformas, en particular, el componente de capacidad analítica, podemos notar cuatro estilos analíticos que se detalladan a continuaciín (Figura 2.3).

Análisis

Entrada Humana



Figura 2.3: Espectro Analitico

Las capacidades analíticas implementadas en organizaciones son a menudo li­mitadas el analisis descriptivo, a traves de reportes basicos y dashboards. Con esto, la pregunta, “Que paso?“ puede ser respondida. Despues de conocer “Que,“ ,lo más probable es que los usuarios también pregunten, “Por que paso?“. Abordar adecuadamente esto requiere mucha mas agilidad y mas capacidades avanzadas de exploración de la información. Despliegues de BI tradicionales tienden a tener huecos en esta area, pero TI por lo general pasa por alto el impacto de esta pro- blematica y contináa impulsando el estándar del proveedor y sus herramientas no aptas-para-proposito. 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 analisis.

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

Mas allá de la capa de datos, vemos la introduction de nuevas herramientas de visualizaciones de datos, y aquá es donde el enfoque del ráapido crecimiento de las herramientas de data discovery se concentran. Pero herramientas tradicionales pueden tambiáen proporcionar mejoras con un mayor enfoque en informes máas comprensivos(como el analisis de varianza), planificacion integrada, dashboards y informes KPI.

A traves del tiempo, con un crecimiento a un alto nivel de madurez de aná­lisis, la organization 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 logicas complejas y reglas de negocio. Ellos incorporan metodos sofisticados que pueden tambien requerir un entendimiento profundo de la estadástica o investigation ope- racional.

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

Redisenar el BI y la Plataforma Analitica Los láderes de BI deben seguir las herramientas fundamentales descritas anteriormente para asá con exito redisenar el BI y la plataforma de analisis. Gartner recomienda la instalacián de una arqui­tectura por niveles compuesta por:

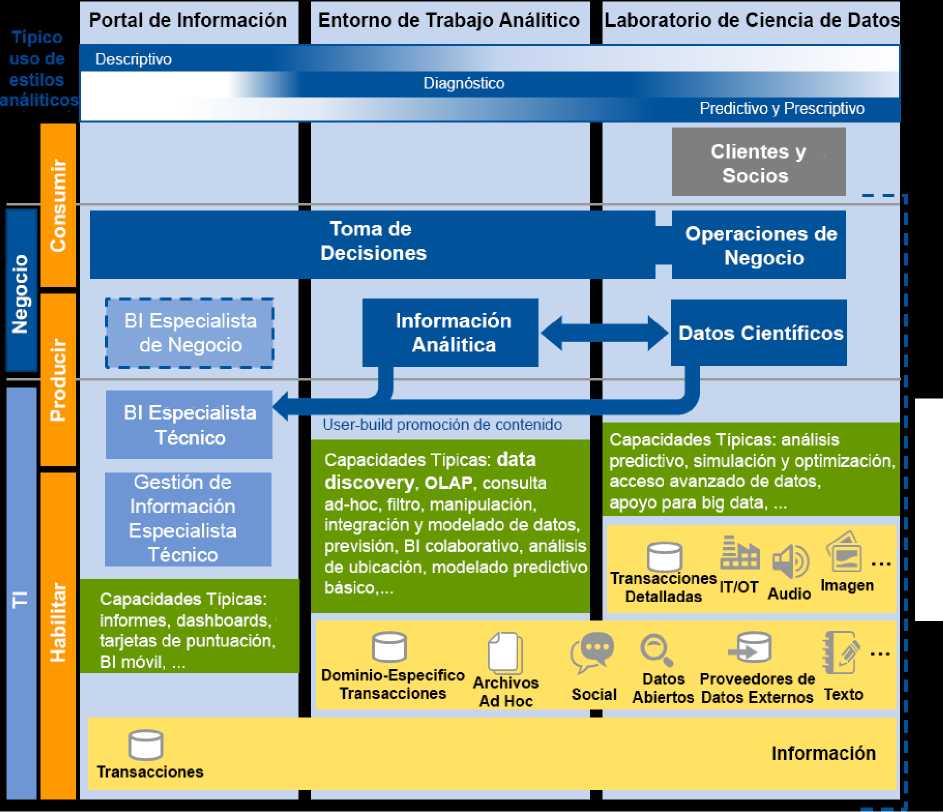
* Portal de Information.
* Workbench Analitico.
* Laboratorio de datos científicos.

En la figura 2.4 se presenta la representation 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.

Figura 2.4: Tipico uso de estilos análiticos

**Gestión - Metadatos**



Para hacer realidad la visián de los tres niveles y ser capaz de maximizar sus fortalezas, los líderes de BI necesitan implementar nuevas capacidades tecnicas para proporcionar nuevos estilos de analisis, mejorar el uso de las herramientas existentes a traves de una mejor integration global, y proporcionar metadatos comunes y gobernanza. Procesos, roles de personas y responsabilidades, son de suma importancia para el íexito.

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

en el framework de análisis de negocios de Gartner.

Vamos a ampliar cada nivel para entender como 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 area de trabajo donde los usuarios de negocio pueden encontrar rópida y facilmente las metricas 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 informacion. Sus productos son el resultado de un proceso de desarrollo formal que abarca que un usuario de negocios establezca requisitos y un especialista tecnico(típicamente de TI, pero cada vez mós de negocio) los implemente.

Capacidades típicas de la plataforma:

* Informes.
* Dashboards.
* Integracion Microsoft Office.
* BI móvil.
* Analisis integrada.

Ejemplo de herramientas y proveedores:

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

1. Workbench analítico

El workbench analítico es el órea 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 agil para explorar informacion y tener acceso a un amplia gama de fuentes de datos, con limitado o ningón apoyo de expertos tecnicos. Los conjuntos de herramientas deberían incluir una herramienta de data discovery (descubrimiento de datos) y un nuómero de otras capacidades para ayudar a los usuarios de negocios a extraer valor de la informacióon de forma autóonoma.

En el espectro de anólisis, el workbench es capaz de proporcionar analisis des­criptivo ,pero por lo general se centraró en el analisis de diagnóstico. En algunos casos - es decir, a traves del uso de herramientas de data discovery mas centrado al analisis. - Puede extender a un nivel bósico de analisis predictivo y ganara el modelado de datos y capacidades analíticas mas avanzadas en el futuro.

Capacidades típicas de la plataforma:

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

- OLAP.

* Mashup de datos y modelado de usuarios de negocios.
* Colaboracion.
* Filtrado y manipulation de datos.

Ejemplo de herramientas y proveedores:

* Tableau Software.
* Qlik.
* Tibco Spotfire.
* SAS Visual Analytics.
* SAP Lumira.
* Oracle Endeca Information 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 area de trabajo donde analisis 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 organization.

Un amplio conjunto de capacidades tecnicas es esperado y a menudo propor­cionado por herramientas especializadas con integration TI mínima, destinada a entregar agilidad y capacidad de responder las preguntas imprevisibles. Esto es porque TI tiende a pasar por alto esta area a favor de inversion en el portal de information. Los usuarios son cualificados y experimentados, a menudo mas que los expertos tecnicos en TI. Sus conjuntos de herramientas incluyen capacidades de data mining, predicciones y otras herramientas de estadísticas y analisis complejos.

Capacidades típicas de la plataforma:

* Acceso avanzado de datos.
* Soporte para fuentes de big data.
* Análisis descriptivo avanzado.
* Analisis predictivo.
* Prediction.
* Optimization.
* Simulation.
* Otros análisis avanzado.
* 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.

|  |  |  |  |
| --- | --- | --- | --- |
| Características | Portal de Información | Entorno de Trabajo Análitico | Laboratorio de  Ciencia de Datos |
| Objectivo Principal | Entrega información estandarizada y confiable para la organización. | Proveer capacidades de exploración de información para una rango extenso de usuarios de negociso. | Permitir la producción de procesos analíticos avanzados. |
| Audiencia | Especialistas técnicos producen el contenido BI; usuarios de negocio(toman decisiones) lo consumen. | Analizar la información de negocio genera contenido para la toma de decisiones, como en el portal de información. | Cientistas de datos generan el contenido para consumido por las operaciones de negocios(como el personal de call center) a través de análitica integrada en aplicaciones de negocio. Puede ser consumido directamente por clientes (Ejemplo: a través de sitios web). |
| Origens de Datos | Información Estructurada desde los almacenes de datos empresariales(EDW) o data mart de dominio específicos. | En general información estructurada desde EDW, data mart de dominio espe­cífico. información generada por usuario(a menudo en  Excel) . aplicaciones de negocio y datos externos y abiertos. Empezar a utilizar fuentes de datos no estructurados. Entradas desde el portal de información. | Información estructurada y no estructurada desde todos los orígenes Internos y externos disponibles. Entradas desde los almacenes analíticos de datos. |
| Confiable vs Ágil | Información Confiable y estructurada, pero estático y poco inflexible. | Interactivo, ágil y personalizable acuerdo a las necesidades del usuario, pero puede mostrar discrepancias en la información generada por el usuario. | Confiable, a fondo y dirigido por datos. Personalizable para responder o resolver un problema simple y poco flexible. |
| Tiempo para Entregar Contenido | Puede llevar dias o meses para desarrollar, pero puede ser consumido en segundos. | Minutos a horas. | Dias a meses. |
| Nivel de Habilidades Requeridas | Habilidades técnicas de intermedio a avanzado para desarrollo de contenido. Sin necesidad de habilidades particulares de BI para el consumo de información. | Básica a anvazada capacidades de manipulación a nivel usuario de negocios. No requiere grandes conocimientos técnicos o estadísticos. | Avanzadas técnicas y habilidades matemáticas. |
| Acceso a  Información  Requerida | Acceso reducido para el consumo de información, de acuerdo al rol y perfil del usuario. | Acceso amplio a multiples data source, de acuerdo a áreas de responsabilidad. | Acceso muy amplio a información, de acuerdo a áreas de responsabilidad. |
| Apoyo TI  Requerido | Elevado desarrollo de contenidos. | Disponibilidad de data source medianos y validación de contenidos. | Baja disponibilidad de data source. |
| Capacidades analíticas típicas Producida | Análisis descriptivo y salidad de análisis prescriptivos y predictivos. | Componentes básicos, descriptivos y diagnósticos de análisis descriptivos( asi como previsión). | Avanzado análisis descriptivo, diagnóstico, predlctlvo y prescrlptlvo. |

Figura 2.5: Características de BI en niveles y Plataforma Analítica

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 analisis de las mejores herramientas del mercado, en un área de conocimiento. A continuation se presenta el Cuadrante Mtigico de Gartner, para herramientas de BI y Analytics.

1. Cuadrante Magico de Gartner

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

**CAPACIDAD PARA EJECUTAR**

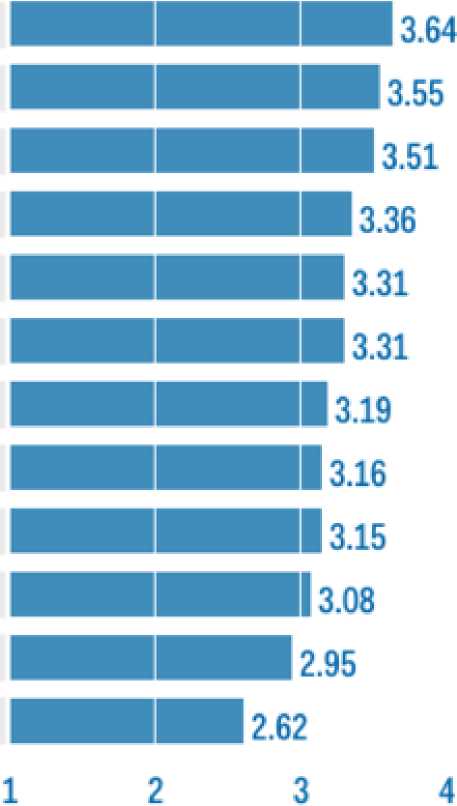
Figura 3.1: Cuadrante Mágico para BI y Plataformas Analíticas



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

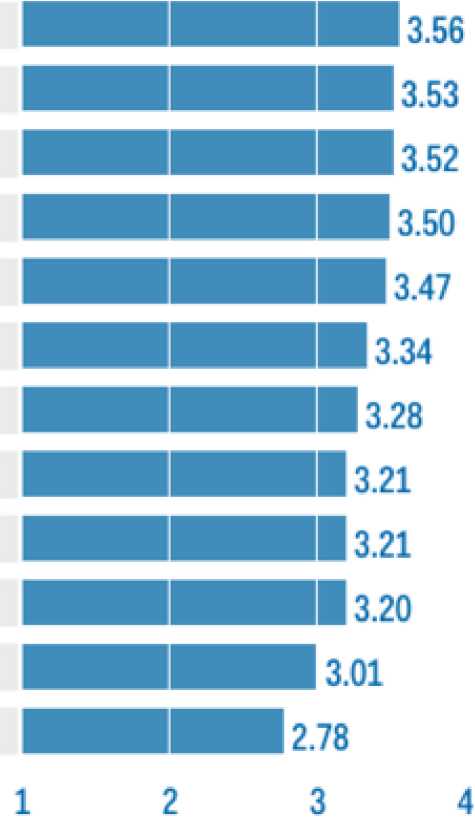
Microsoft

Oracle

Figura 3.2: Puntuaciones de producto o servicio para analisis descentralizado

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

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, ejecucion 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 construction y publicación de dashboards, ademas 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 codificacióon, de esta manera, posibilita a un usuario con escasos conocimientos tóecnicos, poder utilizarlo de igual manera.

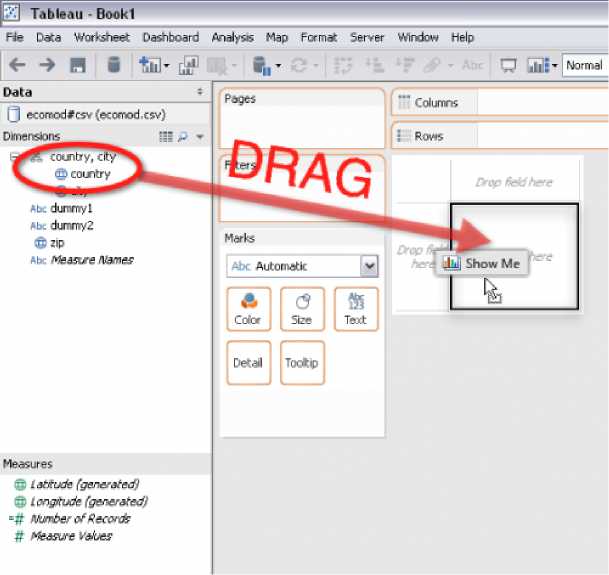


Figura 3.4: Arrastre el campo país para el campo desplegable senalado.

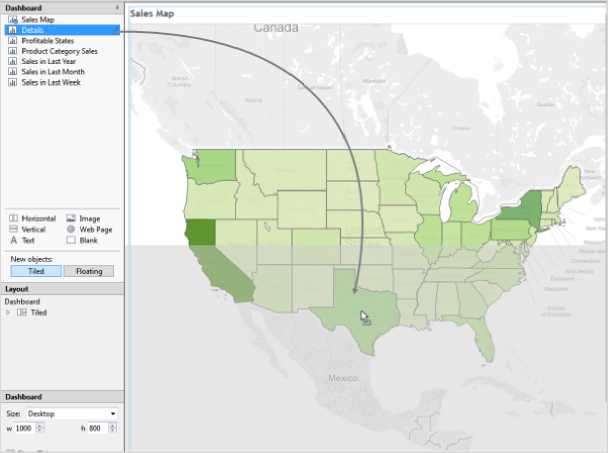


Figura 3.5: Arrastrar hojas de trabajo al dashboard.

De una forma agil 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. Tambien se puede utilizar para realizar proyecciones o tendencias, la cual Tableau nos ofrece de manera automáatica.

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 decision, lo cual podría ocasionar un gran impacto negativo a la organizaciáon. La informaciáon es considerada como uno de los recursos maás im­portantes en una organizacián, y en base a esta informacián, 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 informaciáon de interáes para la organizaciáon.

3.2.1. Datos de la ANDE y de la DGEEC

Se cuenta con datos de consumo de energía electrica, facturaciones, grupos de consumo (residencial, industrial, exportación, comercial, gubernamental y otros), por ano (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 seccion 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 ano 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 formula empleada (ver Formula abajo), donde n es el ano actual y n-1 el ano anterior, PIB indica la cantidad de clientes que posee la ANDE .

tu

PIBn — PIBn-1

PIBn-1

x 100

3.2.2.I. Dashboard - Clientes Facturados vs Crecimiento Poblacional

En el dashboard de la Figura 3.6 son utilizados datos históricos de la po- blacion, 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. Ademas, puede ser muy relevante la informacion de proyeccion del consumo para los próximos años, para elaborar una planificación en la ampliacion de la capacidad de transmisión o distribucion de energía.

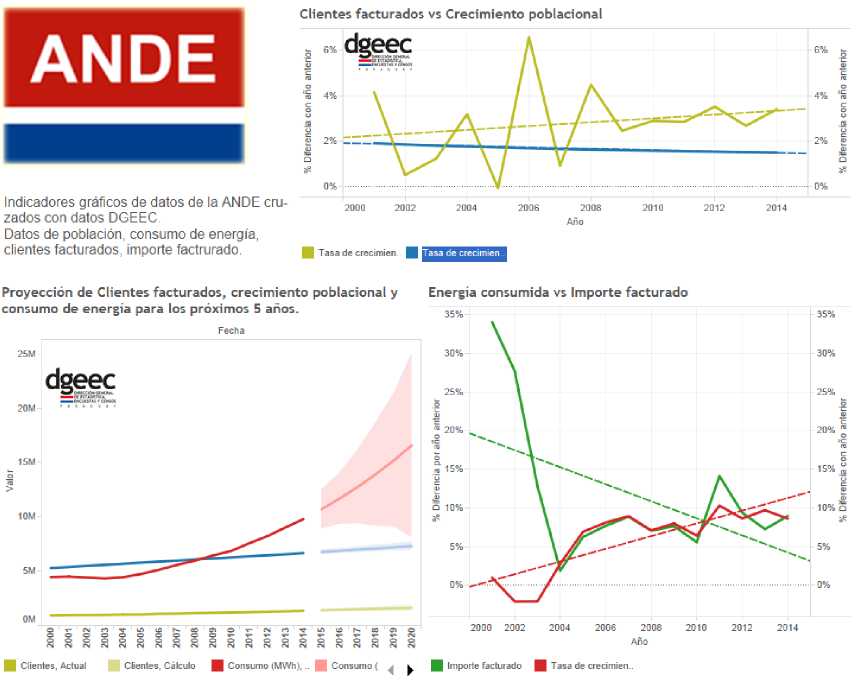
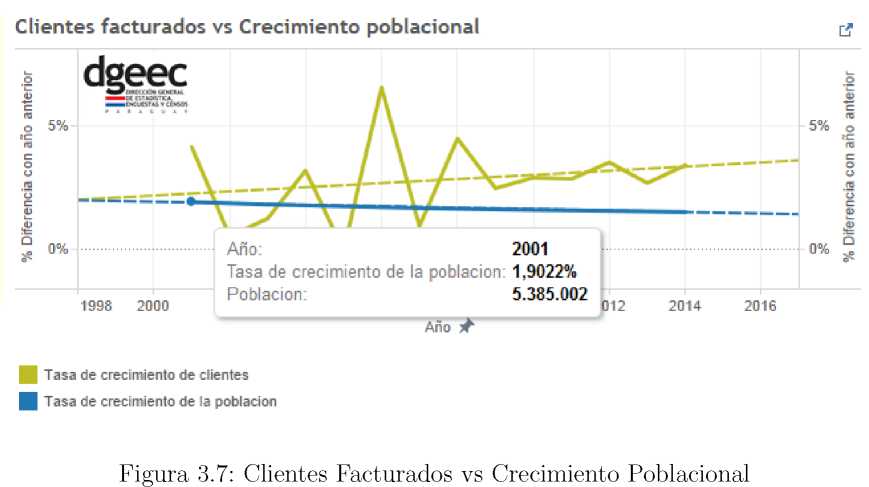


Figura 3.6: Clientes Facturados vs Crecimiento Poblacional

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 poblacion disminuye con el tiempo, pero no así el ritmo de consumo de electri­cidad. No existe una relacion directa entre estas dos tasas por el momento, lo cual puede deberse a factores tales como instalacion de mayor cantidad de dispositivos electricos en residencias, o bien la instalación de mas 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 inversion en la red de electricidad para aquellas zonas donde se registran tendencias mayores al consumo.



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.



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 ano 2001 fue de 959.580, la cual aumento el 4.1 % respecto al ano anterior.

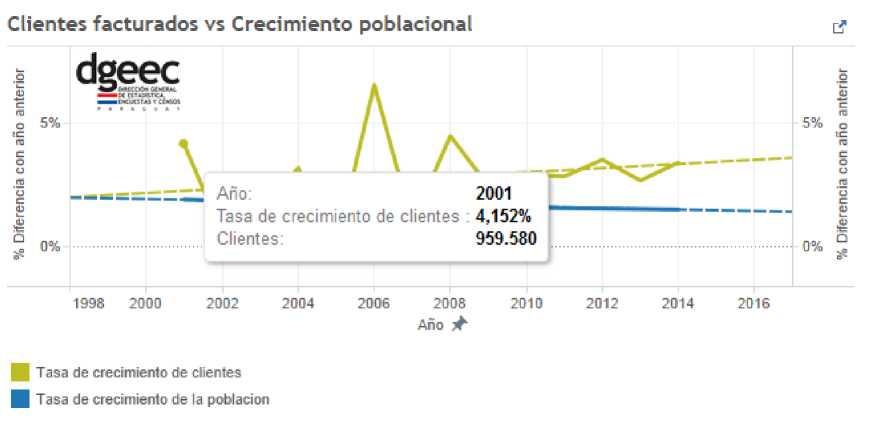
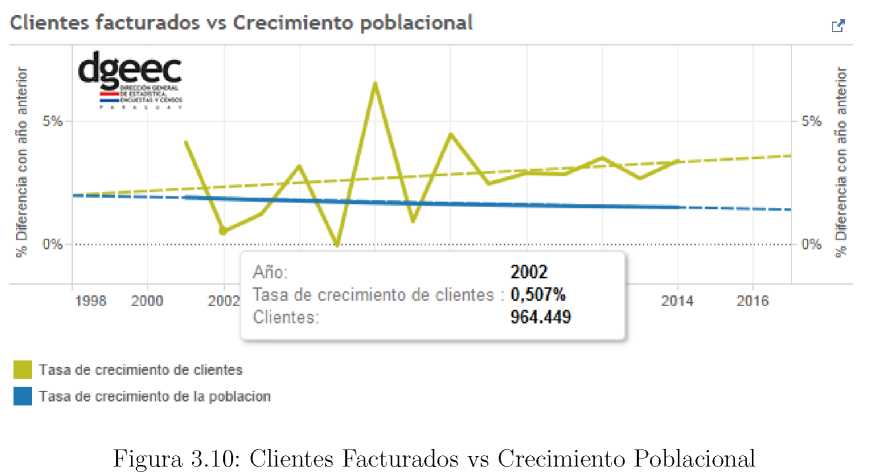


Figura 3.9: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.10, se muestra la cantidad de clientes correspondiente al ano 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 mas que el doble del ano anterior llegando a aumentar 11830 clientes.



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). Ademas, este aumento es mayor en proporcioán al aumento de la poblaciáon y a la cantidad de clientes.

Esto demuestra claramente que el consumo aumenta cada vez mas, indepen­diente que se dáe un aumento en la misma proporciáon de clientes o poblaciáon. 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.

Proyección de Clientes facturados, crecimiento poblacional y rj

consumo de energía para los próximos 5 anos.

Fecha

dgeec

-.'V

15 VI

1QM

Clientes, Actual

Clientes. Calculo

Indicador de pronóstico Cálculo Año de Fecha: 2020

Consumo (MWh): 16.593.716

I

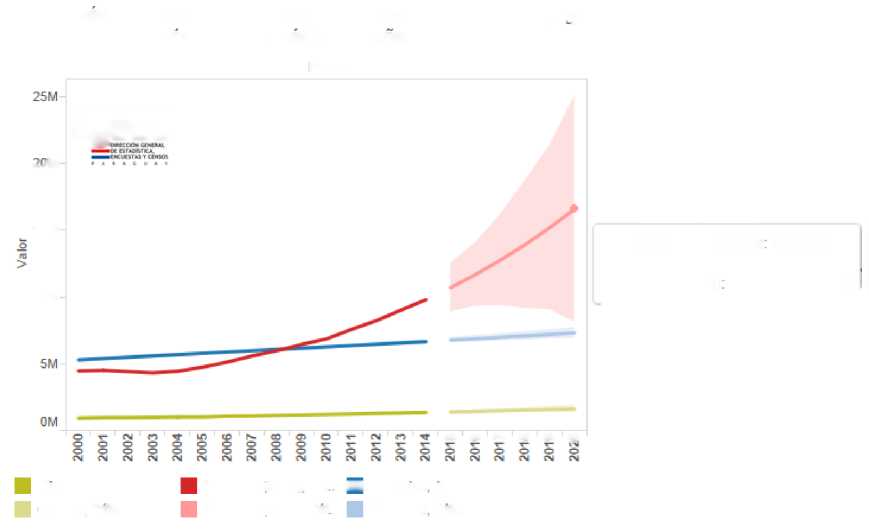
i 10 s a: w p

Consumo iMWh). Act H Población actual

Consumo (MWh), Cal . Población Calculo

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

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 facturacion de la ANDE, en general es proporcional al consumo de energía electrica, excepto en el ano 2011 (Figura 3.13), en la cual el importe facturado fue superior al aumento del consumo de energía. Sin embargo, en el ano 2013 (Figura 3.14) se registra nuevamente un aumento en relacion a años anteriores.



**Energía consumida vs Importe facturado**

tí

30%

**\**

Año 2001

Tasa de crecimiento del importe facturadc 34,03% Importe (GS): 950.600.247

3D%

% Difer&ncia en Imperte (GS)

2D%

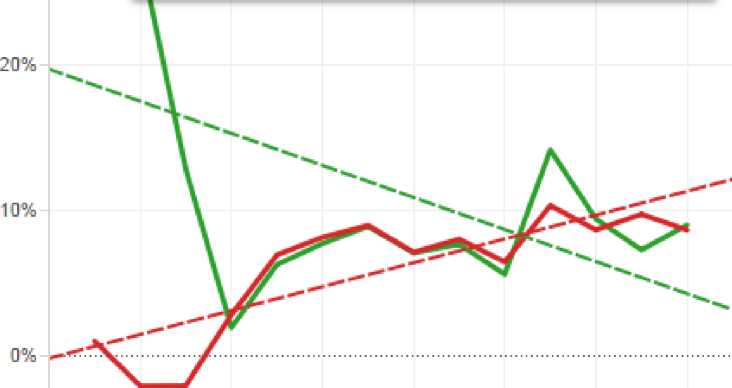
1D%

□ %

2000 2002 2004

2006 2008 2010 Ano

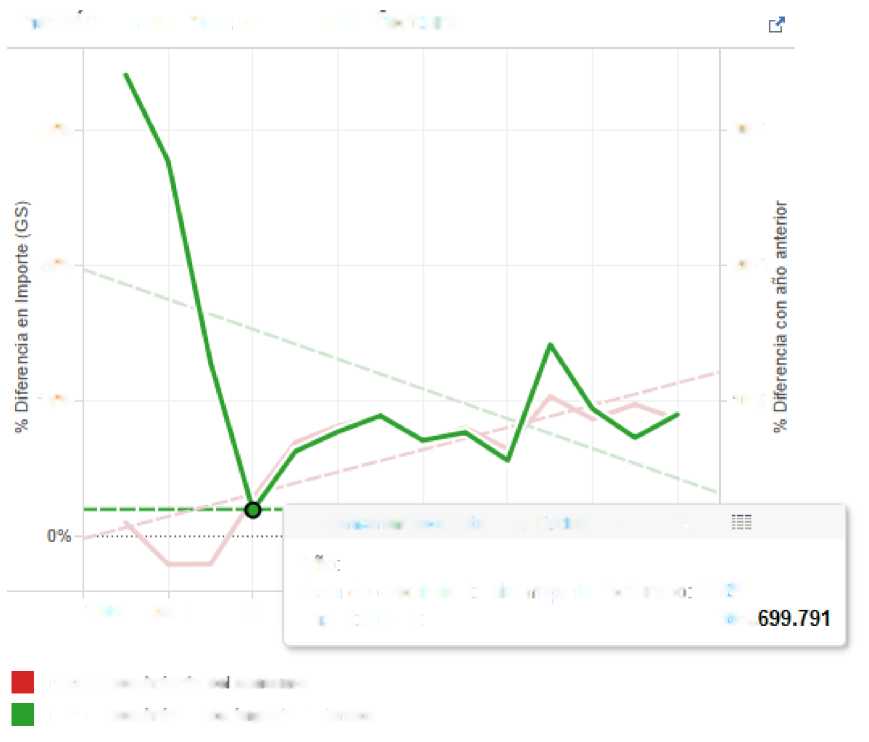
2012 2014



% Difererwiacon ano anterior

Figura 3.13: Energía Consumida vs Importe Facturado

■J Conservar solo esto X Excluir $> % w



Ano 2004

Tasa de crecimiento del imperte facturado 1,89% Importe (GS): 1.394.

2000 2002 2004

23%

13%

13%

**Energía consumida vs Importe facturado**

30%

T asa de crecí merlo de consumo

Tasa de crecimiento del importe facturado

Figura 3.14: Energía Consumida vs Importe Facturado

3.2.2.2. Panel estadístico de consumo de electricidad por sector

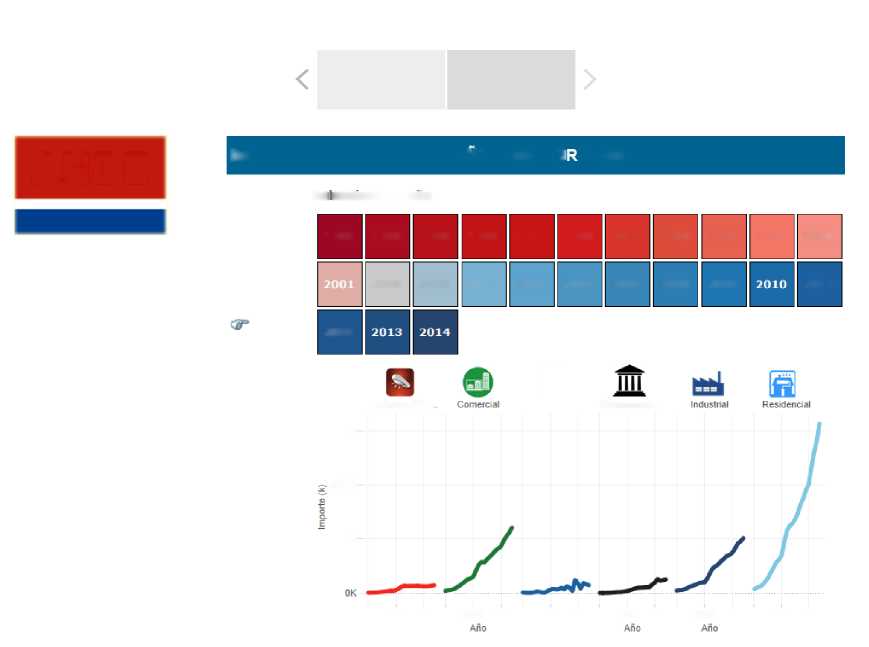
En la figura 3.15 se presenta un gráfico con el consumo de energía históri­ca, que abarca desde el año 1990 hasta 2014. Estos datos están clasificados por los siguientes criterios de la ANDE: Alumbrado Publico, Comercial, Exportacion, Gubernamental, Industrial y Residencial. En este gráfico se puede apreciar que el mayor consumo de energía se encuentra en el sector Residencial. Sin embargo, los valores de Exportacion de energía fue 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 exportacioín.



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

La Figura 3.16 presenta las informaciones de facturación tambien 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.

Datos de la ANDE



Facturación

Consumo

IMPORTE FACTURADO PORANOYSECTO (1990-2014)

**ANDE**

Se eccione un ano

1990

1991

1992

1993

1994

1995

1996

1997

1998

1999

2000

2002

2003

2004

2005

2006

2007

2008

2009

**2011**

1 Mantenga presionado la tecla Ctrl o seleccione con el mouse sobre los años para filtrar.

**2012**

**FACTURACION DE E­LECTRICIDAD.**

Alumbrado Públi.

***&***

Exportación

Gubernamental

En esta vlsualizacion po­demos observar el monto total por año y sector que Ingresa a la institución.

150QK

1G00K

500K

2000 2010 Año

2000 2010

2000 2010 Año

2000 2010

2000 2010

2000 2010 Año

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

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. Tambien se cuenta con un mapa para filtrar por region del país.

Tasa de crecimiento de clientes y consumos por año

ANDE

Distribución por grupo de consumidores

© OpenStreetl.lap contributors

2000 2002

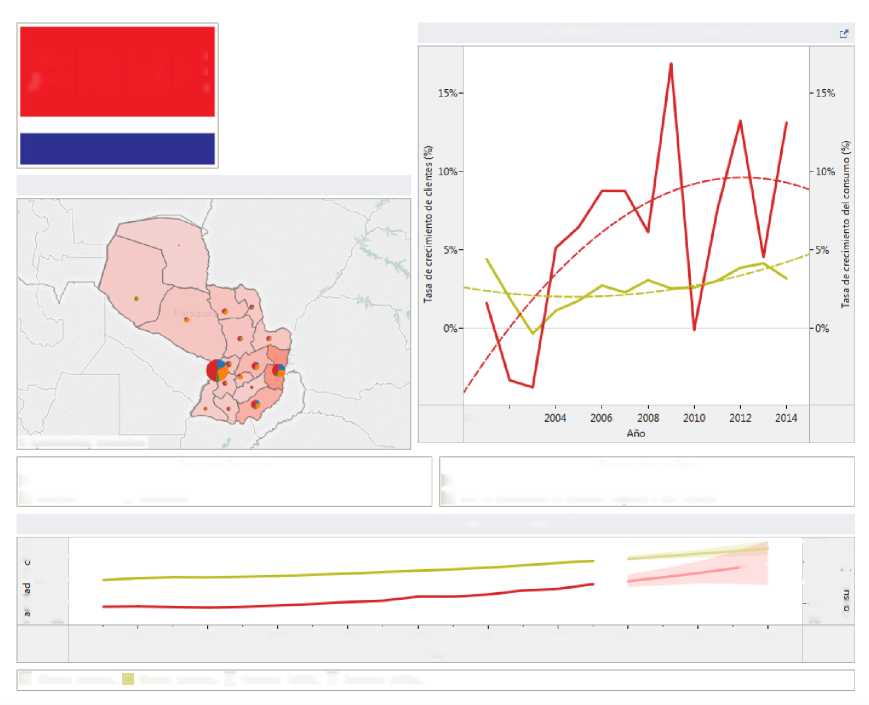
|  |  |  |
| --- | --- | --- |
|  | Grupo de Consumo | Nombre de medidas |
| H Comercial | | Otros | | Tasa de crecimiento de clientes respecto al año anterior |
| | Industrial | | Residencial | | Tasa de crecimiento de consumo respecto al año anterior |

Proyección de dientes y consumos para los próximos 5 años

|  |  |  |
| --- | --- | --- |
| -O  1 1500K- ^ 1000K—  | 500K- u 0K | —1 | "20G £■  ■15G 1 -10G 1  5G =  0G u |
|  |
|  | 2001 2003 2005 2007 2009 2011 2013 2015 2017 2019  Año |  |
| I Clientes (número.. Clientes (número,. | Consumo {MWh).. O Consumo (MWh).. | | |

Figura 3.17: Tasa de crecimiento vs Consumo de energía

En el grafico situado a la derecha del mapa (Figura 3.18), se tiene el consumo y crecimiento de clientes. Para un mejor analisis fue necesario suavizar los datos calculando lineas 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.



**Tasa de crecimiento de dientes y consumos por ano**

L5%-

10%

Q%-

15%

**10**%

Q-¿

**2002**

2004

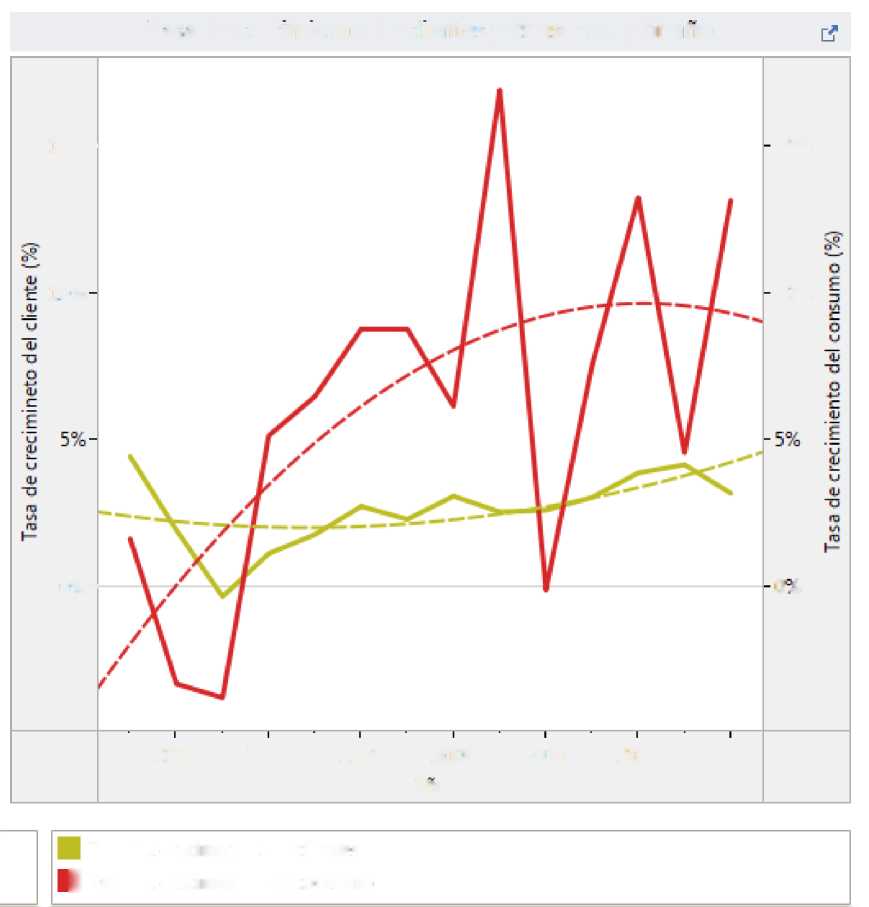
**2005**

**2008**

2010

**2012**

2014



Ano

Tasa de crecimiento de clientes  
| Tasa de crecimiento de consumo

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

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 area de Matematicas Estadísticas. En este gráfico se puede notar que existe una mayor probabilidad que en los práximos 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 anos

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

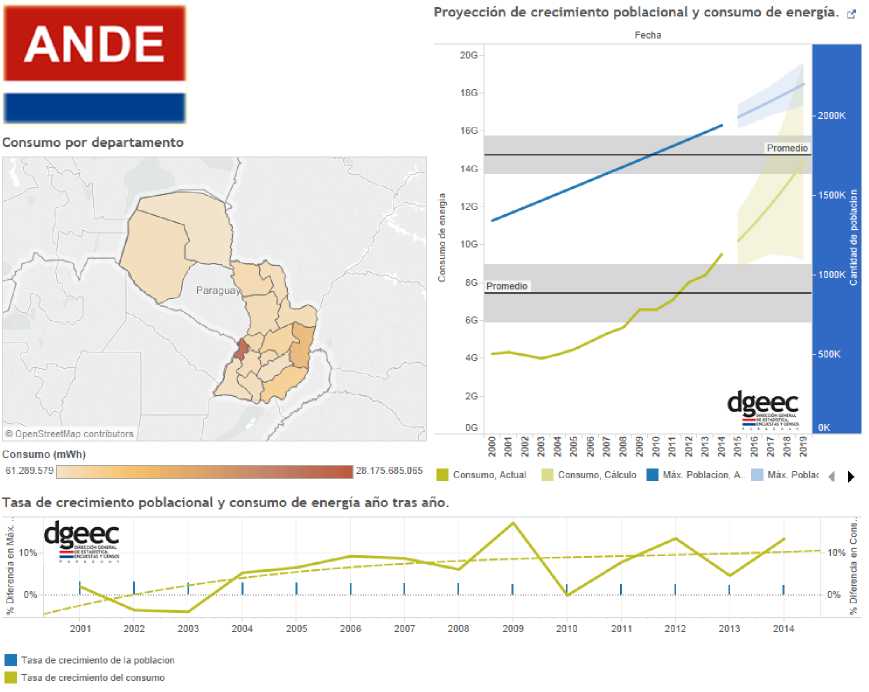
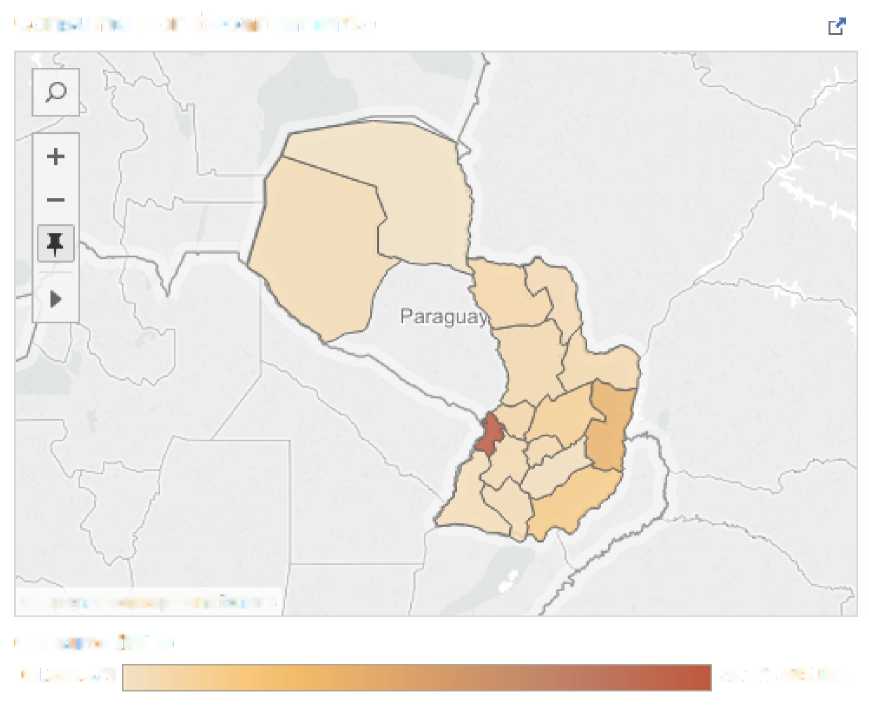


Figura 3.20: Panel de Tasa de crecimiento poblacional y Consumo de energía anual

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 mas oscuro el consumo de energía es mayor, y si el color es mas claro, el consumo es menor. Se puede notar que los departamentos Central y Alto Parana son los que tiene mayor consumo de energía. Este tipo de gráfico es muy util cuando se desea analizar información de forma general y georeferenciada. 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 demías gríaficos tambiíen se actualizan filtrando el departamento seleccionando.



**Consumo por departamento** © QpenStreetMap contributors

**Consumo {kWh)**

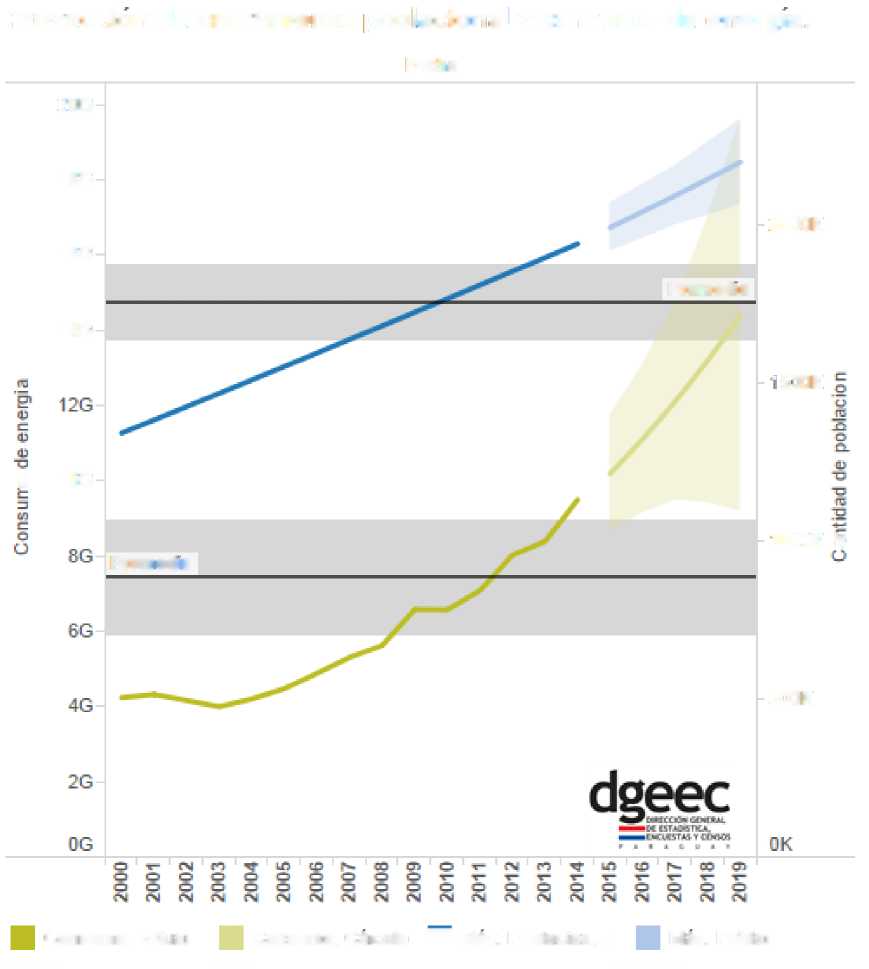
2S.175.6B5.G65

51.289.579

Figura 3.21: Consumo por departamento

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 poblacioín 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 poblacion.

**Proyección de crecimiento pobladora! y consumo de energía.**



Fecha

2CG 13G

**2000K**

16G

Promedio

**14G**

' 5IÜK

S 1DG

**I000K £**

Promedio

500 K

Consumo, Actual Consumo, Calculo ^ Max. Población, A.. Max. Poblar ^ ^

Figura 3.22: Proyección de crecimiento poblacional y consumo ee energía

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

| Tasa de crecimiento de la población H Tasa de crecimiento del consumo

% Diferencia en yáx...

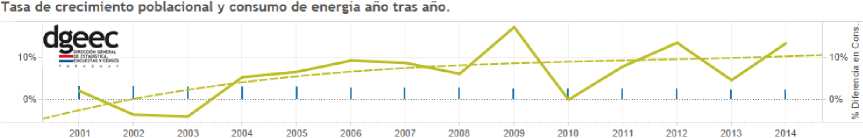


Figura 3.23: Tasa de crecimiento poblacional y consumo de energía años tras años

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 mas arriba para este fin.

1. Enfoque

El enfoque que utilizamos es el cuantitativo, que por lo común, utiliza la reco- leccion y el analisis de datos para contestar preguntas de investigación y probar hipótesis establecidas previamente, y confía en la medicion numerica, 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 poblacion de un país).

1. Técnica e Instrumentos de recolección de datos

La tecnica aplicada en este trabajo en la recoleccion de datos fue la investigation de documentos científicos procedentes de publicaciones de empresas pioneras en Data Discovery y de expertos en el írea.

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 seccion desarrolla el estado del arte en el area de BI y principalmente el de Data Discovery, tecnica aplicada en este trabajo.
* Capítulo 3: En esta seccion se presenta la evaluation tecnica realizada para la selection de la herramienta Tableau. Ademas la aplicacion de tecnicas de Data Discovery a los datos de la ANDE y la DGEEC. Tambien se presentan los productos construidos en este trabajo, para el anaólisis de las informaciones, y las proyecciones realizadas para los próximos anos.
* Capítulo 4: En esta sección se presenta el marco metodologico utilizado para el desarrollo de esta tesis, incluyendo el alcance, el enfoque y la tecnica e instrumentos de recoleccióon de datos.

Para el desarrollo de este trabajo se obtubieron datos de la ANDE y la DGEEC. Con estos datos fue posible aplicar tóecnicas de Data Discovery realizar un analisis de las informaciones, cruzarlos, georeferenciarlos, encontrar lóneas de tendencias y pronosticos de crecimiento a futuro tanto del consumo de energía, de clientes de la ANDE y de la poblacion del país. En cada caso se presenta un analisis que demuestra con gráficos intuitivos que en ciertas ocasiones no existe una relacion 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 energóa exportada.
3. Aunque se tuvo una disminucion 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 (Itaipu) o Argentina (Yacyreta).
4. Segun el pronostico de crecimiento para los próximos años, el consumo de energía tendría un crecimiento mayor al de la cantidad de clientes, y la po- blacion del país.
5. La tasa de crecimiento de la poblacion se mantiene constante durante el tiempo, esto es, la poblacion 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 Inversion en la capacidad de transmision y distribution de electricidad dentro del territorio del país, para los próximos años, ademas de ayudar con indicadores para montar una estrategia de exportation 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 Data Warehouse 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).