



UNIVERSIDAD CATÓLICA
“NUESTRA SEÑORA DE LA ASUNCIÓN”
CAMPUS ALTO PARANÁ
FACULTAD DE CIENCIAS Y TECNOLOGÍA

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

Ariel Hernán Landaida Duarte
Iván Ariel Cáceres Cañete

Hernandarias, abril de 2016

UNIVERSIDAD CATÓLICA
“NUESTRA SEÑORA DE LA ASUNCIÓN”
CAMPUS ALTO PARANÁ
FACULTAD DE CIENCIAS Y TECNOLOGÍA

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

**Ariel Hernán Landaida Duarte
Iván Ariel Cáceres Cañete**

Ricardo Luis Brunelli Montero, Ing.
Tutor

Hernandarias, abril de 2016

Ariel Hernán Landaida Duarte
Iván Ariel Cáceres Cañete

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

Hernandarias, abril de 2016

Landaída 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.

Ariel Hernán Landaida Duarte
Iván Ariel Cáceres Cañete

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. 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:

Hernandarias, abril de 2016

DEDICATORIA

A mis padres, por la oportunidad de existir, por su sacrificio en algún tiempo incomprensido, 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.

Agradecimientos

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.

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 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, Data Warehouse.

ABSTRACT

An investigation was conducted in the area of Business Intelligence (BI) with data from the public institution ANDE in order to estimate potential growth of its structure according to the growth of the population and can have a better understanding of the data and facilitate decision-making.

Keywords: Data Discovery, Business Intelligence, Data Warehouse.

ÍNDICE DE FIGURAS

| | | |
|-------|--|----|
| 2.1. | Etapas de BI como fuentes de datos de calidad | 6 |
| 2.2. | Framework Analítico de Negocios Gartner | 18 |
| 2.3. | Espectro Analítico | 19 |
| 2.4. | Tipico uso de estilos analíticos | 21 |
| 2.5. | Características de BI en niveles y Plataforma Analítica . . | 25 |
| 3.1. | Cuadrante Mágico para BI y Plataformas Analíticas | 28 |
| 3.2. | Puntuaciones de producto o servicio para análisis descen- tralizado | 29 |
| 3.3. | Puntuaciones de producto o servicio guiados por Data Dis- covery | 30 |
| 3.4. | Arrastre el campo país para el campo desplegable señalado. | 31 |
| 3.5. | Arrastrar hojas de trabajo al dashboard. | 32 |
| 3.6. | | 32 |
| 3.7. | Formúla para hallar tasa de crecimiento. | 33 |
| 3.8. | Clientes Facturados vs Crecimiento Poblacional | 34 |
| 3.9. | Clientes Facturados vs Crecimiento Poblacional | 35 |
| 3.10. | Clientes Facturados vs Crecimiento Poblacional | 35 |
| 3.11. | Clientes Facturados vs Crecimiento Poblacional | 36 |
| 3.12. | Clientes Facturados vs Crecimiento Poblacional | 37 |
| 3.13. | Clientes Facturados vs Crecimiento Poblacional | 37 |
| 3.14. | Proyección de Clientes Facturados Crecimiento Poblacional Y Consumo De Energía para los Próximos 5 Años | 38 |
| 3.15. | Energía Consumida vs Importe Facturado | 39 |
| 3.16. | Energía Consumida vs Importe Facturado | 40 |
| 3.17. | Estadística de consumo de electricidad por sector(1990-2014) | 41 |
| 3.18. | Importe facturado por año y sector(1990-2014) | 42 |
| 3.19. | Tasa de crecimiento vs Consumo de energía | 43 |
| 3.20. | Tasa de crecimiento vs Consumo de energía, filtrado por el departamento Alto Paraná | 44 |
| 3.21. | Tasa de crecimiento vs Consumo de energía | 45 |
| 3.22. | Proyección de clientes y consumos para los próximos 5 años | 46 |

| | | |
|-------|---|----|
| 3.23. | Dashboard - Tasa de crecimiento poblacional vs consumo de energía año tras año | 47 |
| 3.24. | Consumo por departamento | 48 |
| 3.25. | Proyección de crecimiento poblacional y consumo ee energía | 49 |
| 3.26. | Tasa de crecimiento poblacional y consumo ee energía años tras años | 50 |

LISTA DE TABLAS

| | |
|--|----|
| 2.1. Técnicas actuales de BI | 14 |
|--|----|

ÍNDICE

| | |
|--|-----------|
| 1. CAPITULO 1 | 1 |
| 1.1. Introducción | 1 |
| 1.2. Planteamiento del problema | 1 |
| 1.3. Objetivos | 2 |
| 1.3.1. Objetivo General | 2 |
| 1.3.2. Objetivos Específicos | 2 |
| 1.4. Justificación | 2 |
| 2. CAPITULO 2 | 5 |
| 2.1. MARCO TEÓRICO | 5 |
| 2.1.1. Business Intelligence | 5 |
| 2.1.2. Data Discovery Analysis | 15 |
| 2.1.3. Framework de Análisis de Negocio de Gartner | 17 |
| 3. CAPITULO 3 | 27 |
| 3.1. Selección de la herramienta para Data Discovery | 27 |
| 3.1.1. Cuadrante Mágico de Gartner | 27 |
| 3.2. Aplicación de Data Discovery a datos de instituciones del Estado . | 32 |
| 3.2.1. Dashboard - Clientes Facturados vs Crecimiento Poblacional | 34 |
| 3.2.2. Dashboard - Estadística de consumo de electricidad por sector | 41 |
| 3.2.3. Dashboard - Tasa de crecimiento vs Consumo de energía . . | 43 |
| 3.2.4. Dashboard - Tasa de crecimiento poblacional vs Consumo de energía año tras año | 47 |
| 4. CAPITULO 4 | 51 |
| 4.1. Marco Metodológico | 51 |
| 4.1.1. Alcance | 51 |
| 4.1.2. Enfoque | 51 |
| 4.1.3. Técnica e Instrumentos de recolección de datos | 51 |
| 5. CAPITULO 5 | 53 |
| 5.1. Conclusiones y Trabajos futuros | 53 |

| | |
|-----------------------|----|
| Referencias | 55 |
|-----------------------|----|

CAPITULO 1

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. 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 apoyó 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.

1.2. 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 suficiente 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 los datos suficientes pero no consigue analizarlos o procesarlos por falta de conocimiento del negocio, además de desconocimiento técnico de las herramientas.

1.3. Objetivos

1.3.1. 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.

1.3.2. 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.

1.4. 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

2.1. MARCO TEÓRICO

2.1.1. 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 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 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, decision support system)(“Business Intelligence — Wikipedia, The Free Encyclopedia”, 2015).

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).

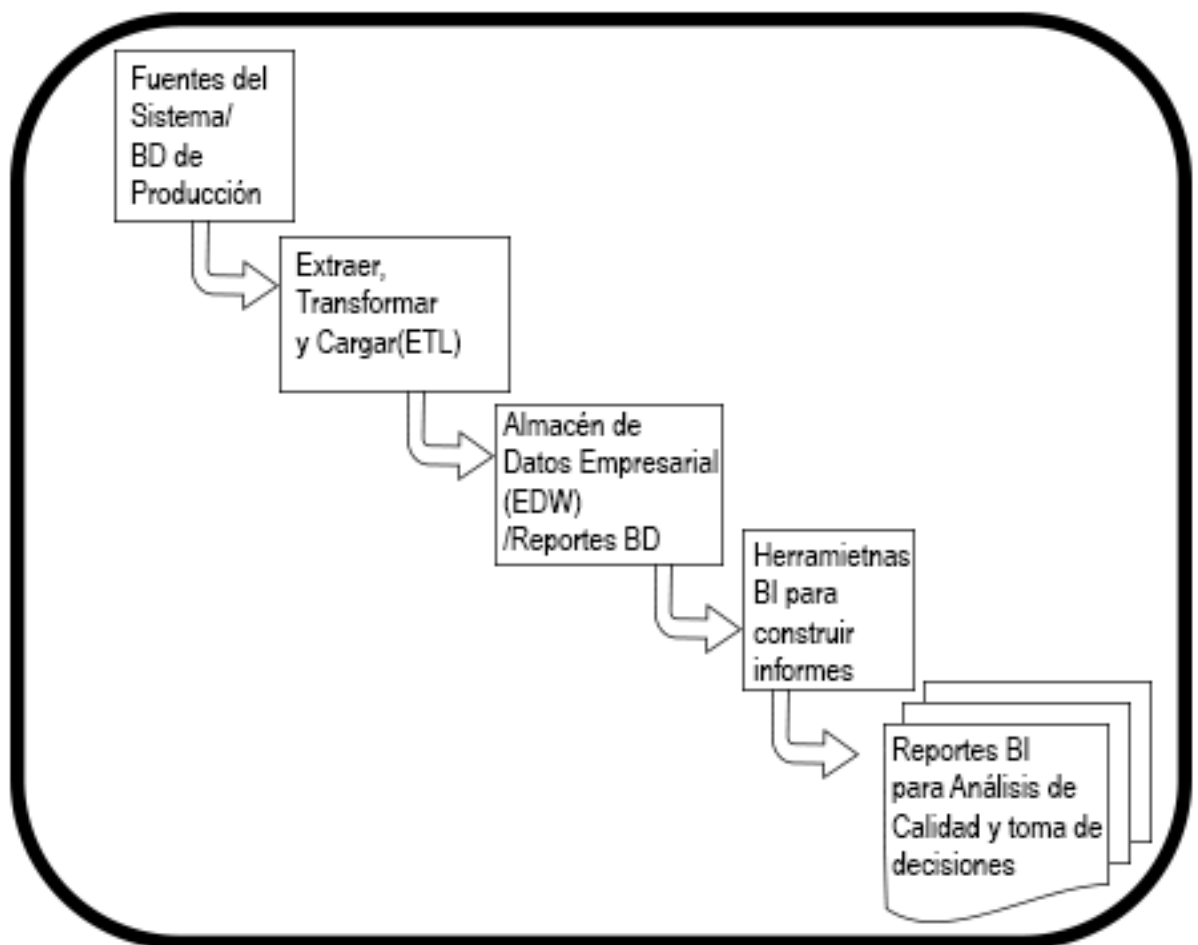


Figura 2.1: Etapas de BI como fuentes de datos de calidad

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 di-

rección de una empresa y el área de TI (tecnología de la información) deberían trabajar juntos para asegurar datos de alta calidad (Eckerson, 2009).

2.1.1.1. Componentes de BI

2.1.1.1.1. 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.(Ranjan, 2009)

Las técnicas y herramientas OLAP pueden ser usados para trabajar con datawarehouse o con data marts (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.(Ranjan, 2009)

2.1.1.1.2. 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.(Ranjan, 2009)

2.1.1.1.3. BI en tiempo real: permite la distribución de métricas en tiempo real a través de emails, sistemas de mensajería instantánea y/o pantallas interactivas.

2.1.1.1.4. 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.(Ranjan, 2009)

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 tácticas de un asunto específico. 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, desde las empresas de investigación de mercados o desde internet, datos no estructurados de redes sociales o información de un datawarehouse existente. Estas fuentes pueden almacenarse en bases de datos relacionales o cualquier otra estructura de datos de sistemas transaccionales. (Ranjan, 2009). También pueden residir en muchas plataformas diferentes, tales como tablas, hojas de cálculo o información no estructurada, como archivos de texto plano, imágenes y otras informaciones multimedia.

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 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 área. Similar al datawarehouse, los data marts 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 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 decisión, éstas 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)

2.1.1.2. Beneficios de BI

BI provee muchos beneficios a las compañías 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 más 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 continuación 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 cuales 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.

2.1.1.3. 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 data warehouse 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 data warehouse.
- Volumen de datos (capacidad).
- Cuánto 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 típicamente 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 Data warehouses.
- 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 indicators) 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 sólo 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.

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 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 aplicación, así 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) introdujo 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 data warehouses y data marts) son BI, si y sólo 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 back-end, una interfaz de usuario en el front-end, 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 multidimensionales, 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 multidimensionales 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).

| TÉCNICAS | DESCRIPCIÓN |
|--|---|
| 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 relación 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 secuencial | 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 visualización | 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

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.

2.1.2. 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. 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 más 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 especializadas. 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. (“Data discovery — Wikipedia, The Free Encyclopedia”, 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 (BusinessObjects), 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 cobró 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 aún 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 sólo 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 desarrollarán 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.

2.1.3. 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, así como una percepción errónea de sus objetivos y cómo 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 data warehouse 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 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 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

El framework de análisis de Gartner identifica las personas, los procesos y componentes de la plataforma que apoyan la transformación de la información en un mejor rendimiento de la organización. El uso de esta herramienta está hecha 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 necesitará 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 detallan a continuación (Figura 2.3)



Figura 2.3: Espectro Analítico

Las capacidades analíticas implementadas en organizaciones son a menudo limitadas al análisis descriptivo, a través de reportes básicos y dashboards. Con esto,

la pregunta, “¿Qué paso?” puede ser respondida. Después de conocer “¿Qué?”, lo más probable es que los usuarios también pregunten, “¿Por qué pasó?”. 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 y 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 significativo 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 así 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.

Esta es la representación de BI en niveles y la plataforma analítica(ver Figura 2.4) “ utilizarlo como una guía genérica que puede ser ajustado de acuerdo a las características específicas de la organización.

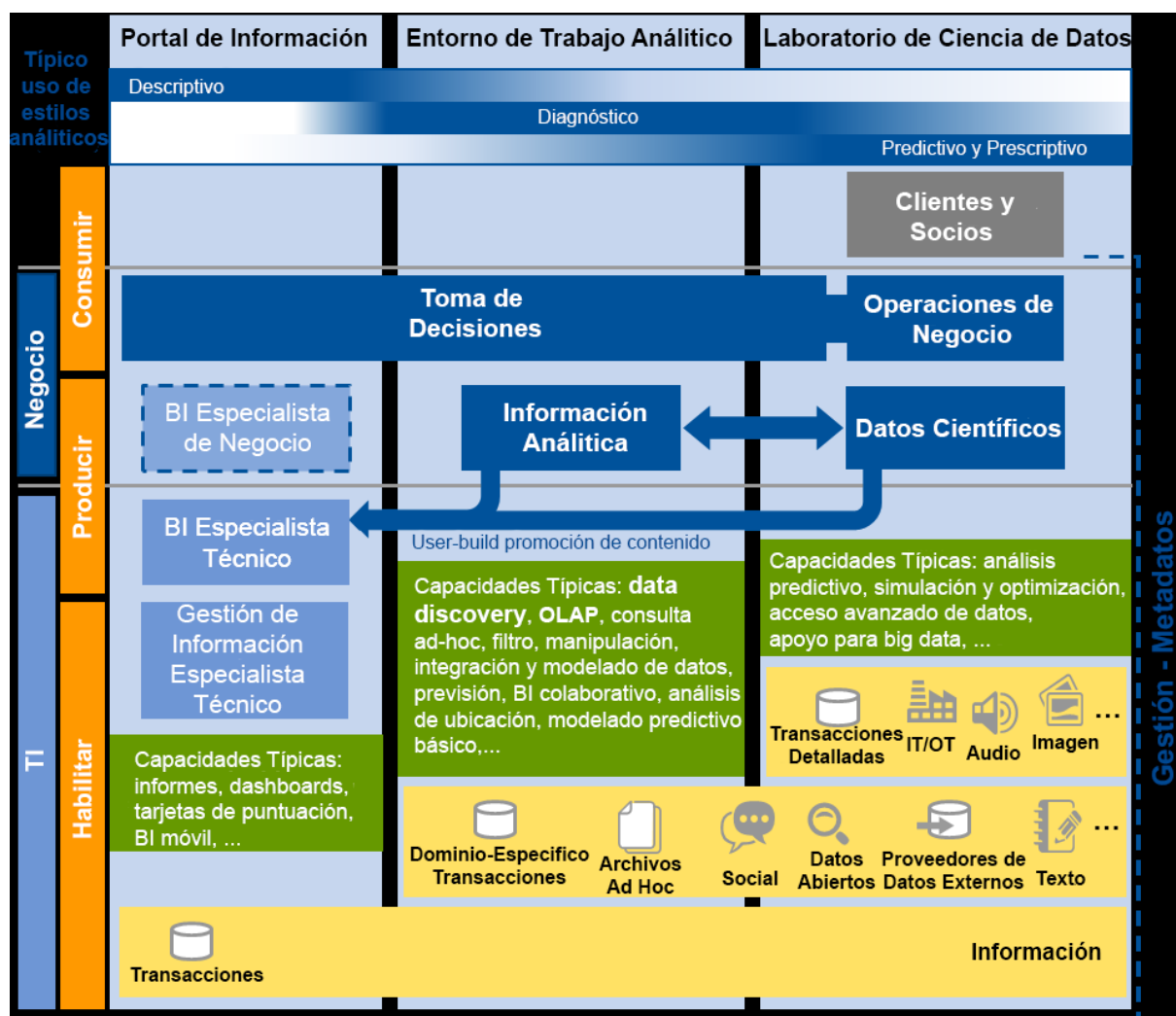


Figura 2.4: Tipico uso de estilos analíticos

Para hacer realidad la visión de los tres niveles (de acuerdo con el Pace layer model) y ser capaz de maximizar sus fortalezas, los líderes de BI necesitan implementar nuevas capacidades técnicas para proporcionar estilos de análisis perdidos, 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, a pesar de no ser detallados

en esta nota, son de suma importancia para el éxito. 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.

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 integrada.

Ejemplo de herramientas y proveedores:

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

Workbench analítico. Seguir de cerca las características de los sistemas de diferenciación desde el Pace Layer Model. 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 un 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 ganará el modelado de datos y capacidades analíticas más avanzadas en el futuro.

Capacidades típicas de la plataforma:

- Data discovery.
- Ad hoc informes/consultas.
- 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 Information Discovery.
- MicroStrategy Visual Insight.
- Alteryx.
- Microsoft SQL Server Analysis Services and Power BI.

Laboratorio de datos científicos.

Seguir de cerca las características de los sistemas de innovación de la Pace Layer Model. 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.
- Predicción.
- Optimización.
- Simulación.
- Otros análisis avanzado.
- Aunque no capacidades BI, Hadoop y otras 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 de las tres capas. Los líderes de BI deberían intentar entender los huecos en su BI e implementación analítica actual y cambiar sus estrategias en consecuencia.

| Características | Portal de Información | Entorno de Trabajo Analítico | Laboratorio de Ciencia de Datos |
|---|---|---|---|
| Objetivo Principal | Entrega información estandarizada y confiable para la organización. | Proveer capacidades de exploración de información para una rango extenso de usuarios de negocio. | 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 analítica integrada en aplicaciones de negocio. Puede ser consumido directamente por clientes (Ejemplo: a través de sitios web). |
| Origenes 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 días o meses para desarrollar, pero puede ser consumido en segundos. | Minutos a horas. | Días 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 avanzada 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 múltiples 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 salud de análisis prescriptivos y predictivos. | Componentes básicos, descriptivos y diagnósticos de análisis descriptivos(así como previsión). | Avanzado análisis descriptivo, diagnóstico, predictivo y prescriptivo. |

Figura 2.5: Características de BI en niveles y Plataforma Analítica

CAPITULO 3

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.

3.1.1. Cuadrante Mágico de Gartner



Figura 3.1: Cuadrante Mágico para BI y Plataformas Analíticas

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

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.

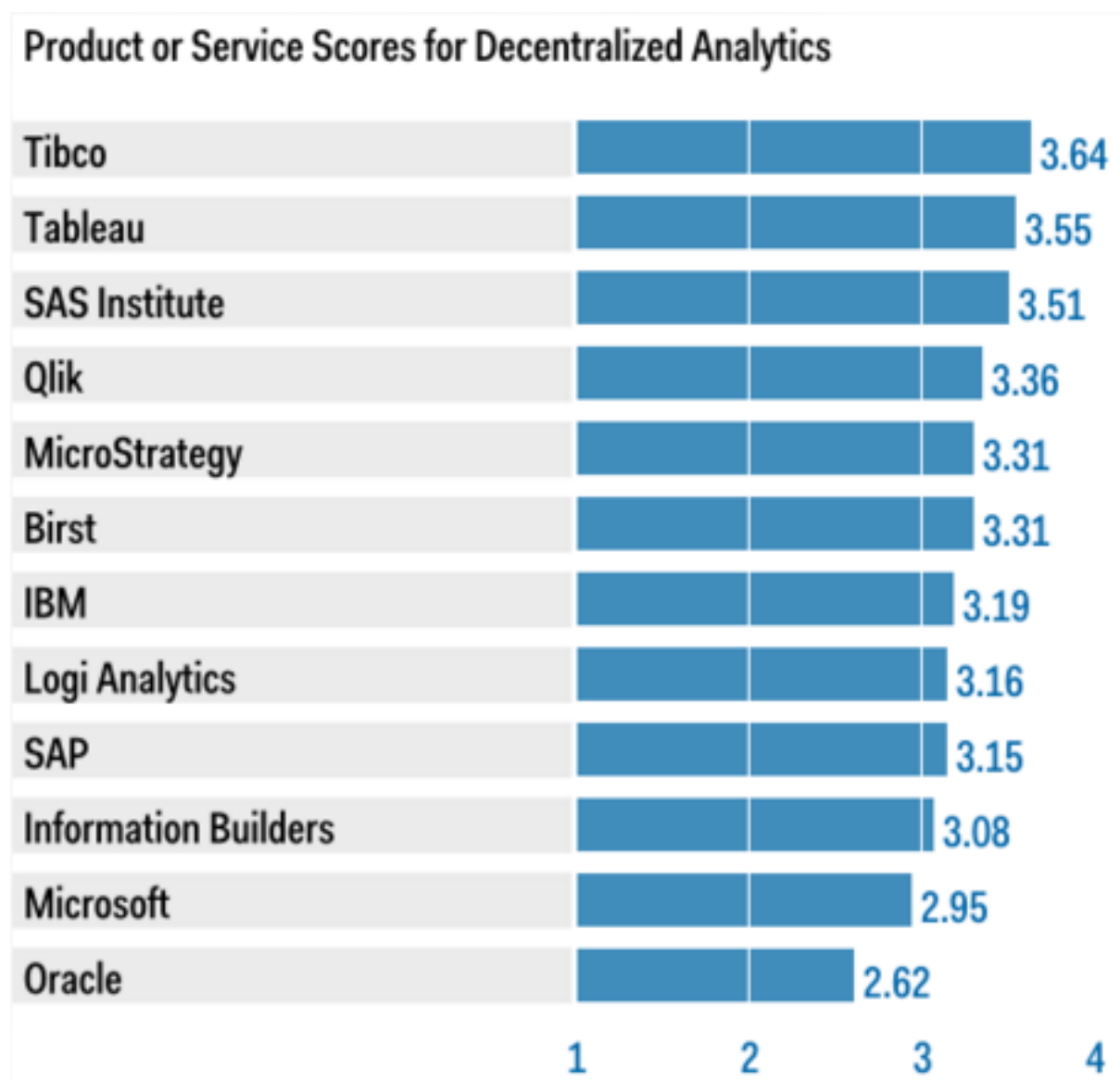


Figura 3.2: Puntuaciones de producto o servicio para análisis descentralizado

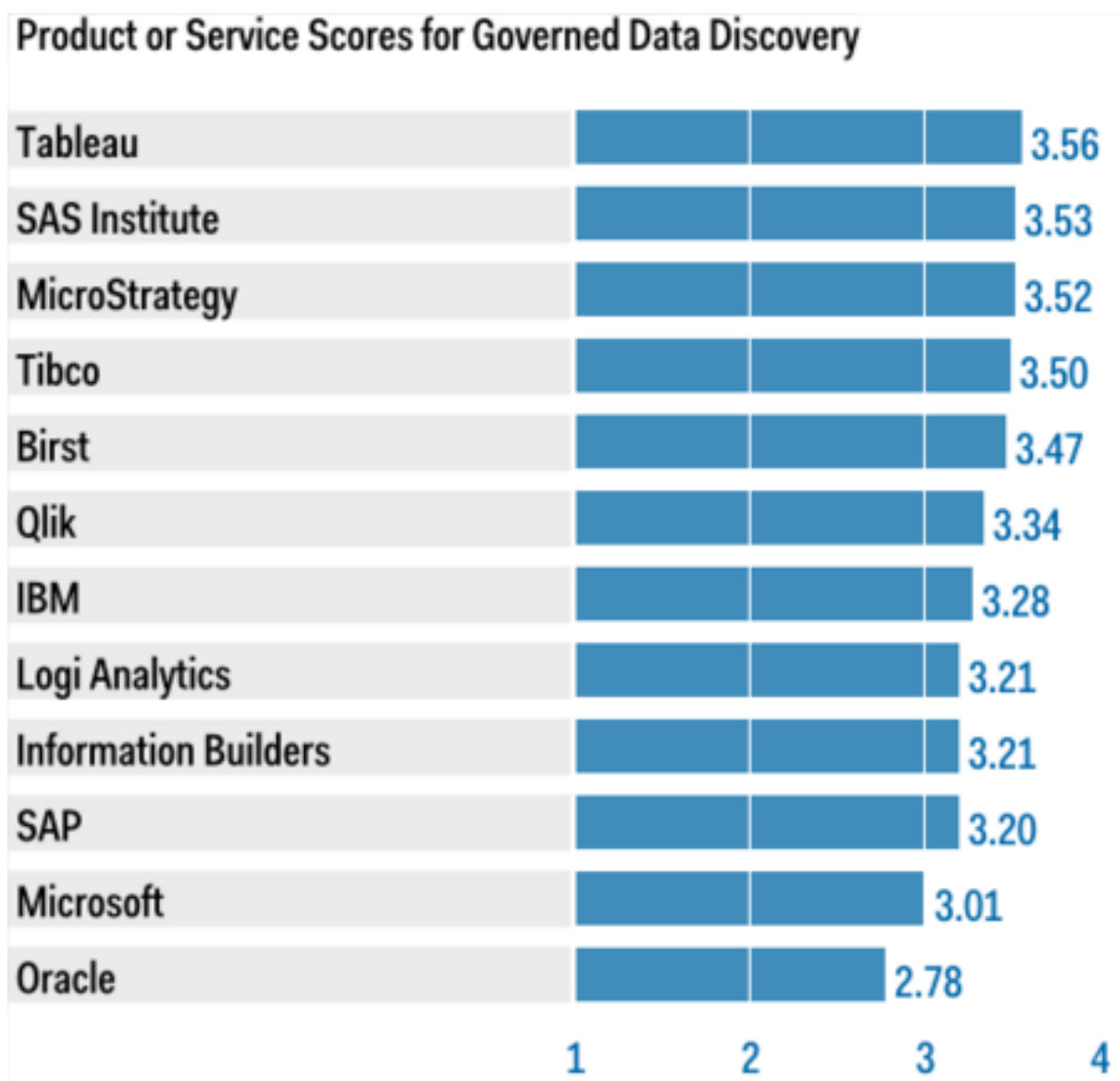


Figura 3.3: Puntuaciones de producto o servicio guiados por Data Discovery

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, ejecución de ventas, marketing, experiencia del cliente) en el eje de líderes del cuadrante. Esta herramienta fue la que mejor se adecuó a las necesidades del trabajo de Tesis, dado que cuenta con una versión pública 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.

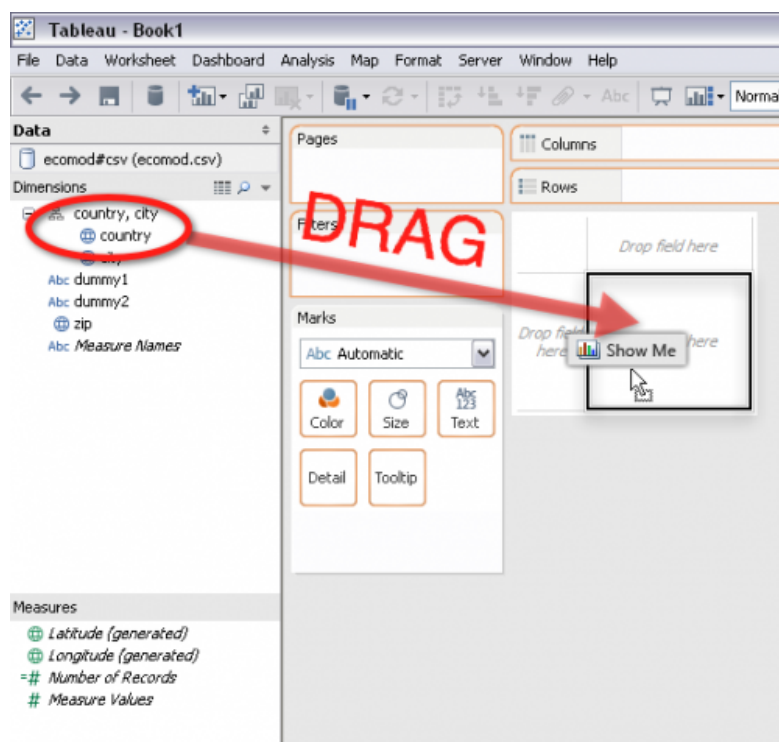


Figura 3.4: Arrastre el campo país para el campo desplegable señalado.

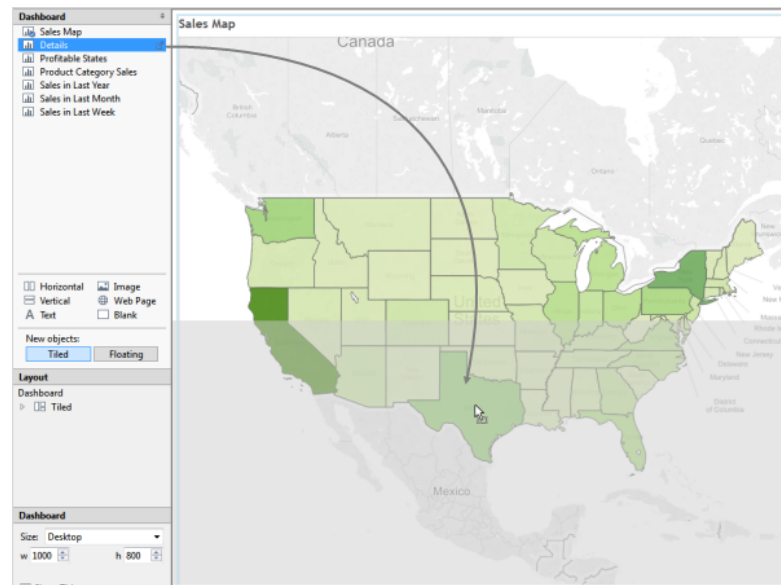


Figura 3.5: Arrastrar hojas de trabajo al dashboard.

En pocos pasos el usuario puede conectarse a diversas fuentes de datos y crear dashboards interactivos, conectando entre sí los diferentes componentes (tipo de gráfico) que nos proporciona la herramienta, si así se desea. La herramienta permite utilizar un componente como filtro para otro, siendo o no de la misma fuente de datos siempre y cuando los datos coincidan en los diversos conjuntos. En las empresas generalmente la utilizan para entender rápidamente distintos aspectos de sus negocios. También podrían utilizar para realizar proyecciones o tendencias, la cual estos cálculos, tableau nos ofrece de manera automática.

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

En Paraguay, e inclusive mundialmente, la mayoría de las empresas no logran comprender 100 % los datos que generan. La consecuencia de no comprender esos datos puede conllevar a tomar una decisión equivocada y esa decisión puede ocasionar un gran impacto negativo a la organización a lo largo del tiempo. La información es considerada como uno de los recursos más importantes en una empresa, porque en base a esto se obtiene conocimiento que podría ayudar a obtener mejores resultados.



Figura 3.6:

Aplicando la técnica de data discovery podríamos detectar irregularidades y oportunidades futuras. En este trabajo analizaremos los datos de la ANDE y de la DGEEC, relacionando ambos conjuntos de datos, con el objetivo de obtener información de interés para la institución.

3.2.0.1. 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), departamento y distrito. Estos datos fueron solicitados formalmente a la institución a través de la Facultad de Ciencias y tecnología de la Universidad Católica, la cual tuvimos una respuesta favorable para proceder.

3.2.0.2. Dashboard de control / monitoramiento

En esta sección se muestra 4 ejemplos de paneles informativos, resultantes del relacionamiento de ambos conjunto de datos. Una de las técnicas utilizada para medir el crecimiento es la “tasa de crecimiento”, la cual se calcula el porcentaje de crecimiento que hubo por cada año (Ej: Si al cerrar el año 2014, la cantidad de clientes llegó a 1.000.000 y en el año 2015 aumentó 100.000, 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 respecto al año anterior). Suponiendo que en el año 2016 la ANDE cierra con un total de 1.000.000 de clientes, su crecimiento sería 10 % menor al año anterior. La fórmula empleada (ver Figura 3.7), donde “n” es el año actual y “n-1” el año anterior, PIB es una variable que indica, en el caso de nuestra comparación, la cantidad de clientes que posee la ANDE .

$$t_n = \frac{PIB_n - PIB_{n-1}}{PIB_{n-1}} \times 100$$

Figura 3.7: Fórmula para hallar tasa de crecimiento.

3.2.1. Dashboard - Clientes Facturados vs Crecimiento Poblacional

En el dashboard (ver Figura 3.8) se utilizan datos históricos de la población que se obtuvo del conjunto de datos de la *Dirección General de Estadísticas, Encuestas y Censo* (DGEEC) y datos de clientes, consumo y facturaciones del conjunto de datos de la *Administración Nacional de Electricidad* (ANDE).

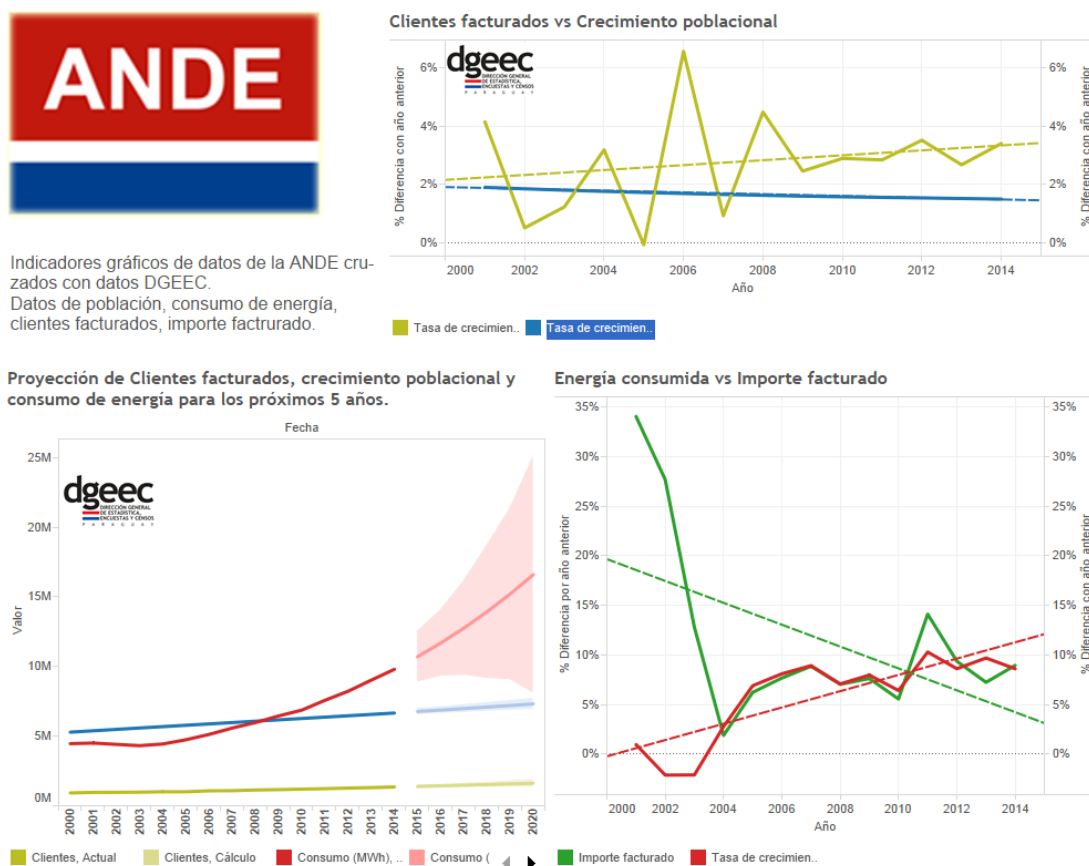


Figura 3.8: Clientes Facturados vs Crecimiento Poblacional

La figura 3.9 expone el porcentaje del crecimiento anual de los clientes en comparación con el de la población. Podemos observar que la línea azul, la cual pertenece a la tasa de crecimiento de la población, fue bajando con el tiempo. Esto no quiere decir que la población fue disminuyendo, sino que cada año el porcentaje de aumento es menor. En el año 2001 la población cerró con un total de 5.385.002.

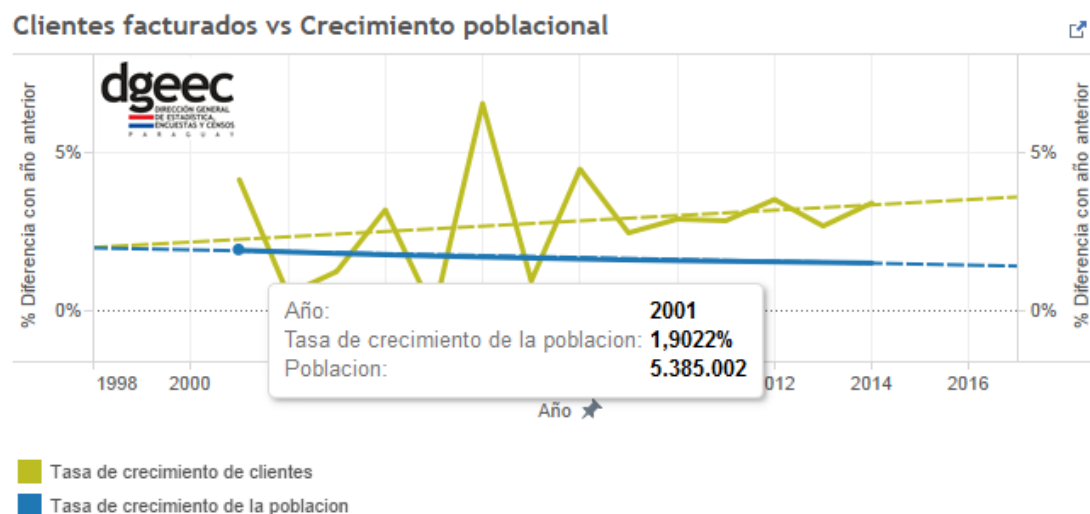


Figura 3.9: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.10 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.

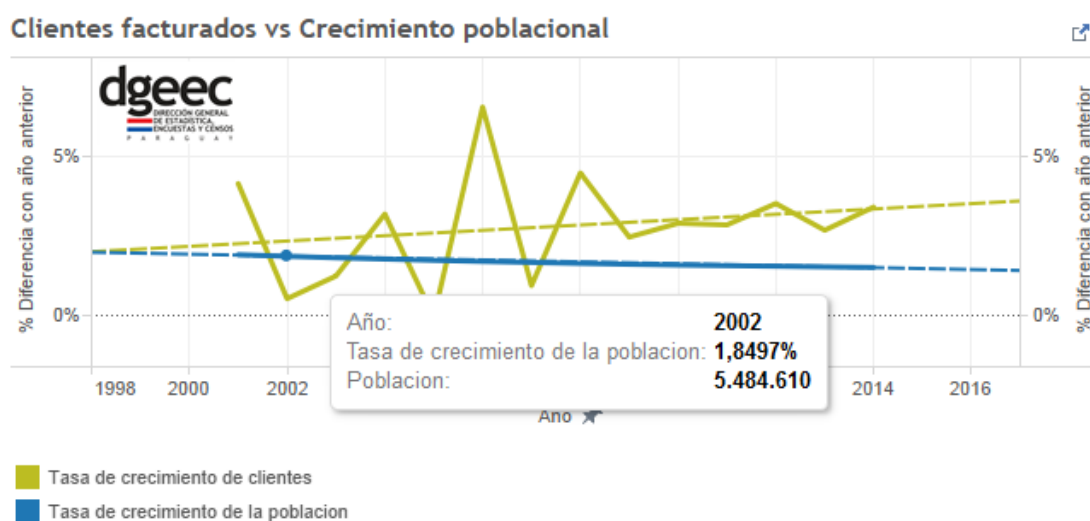


Figura 3.10: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.11 se muestran valores que representan el porcentaje del crecimiento de los clientes de la ANDE. Como podemos ver, hay años 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 aumentó el 4.1 % respecto al año anterior.



Figura 3.11: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.12, 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 3.13, 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.

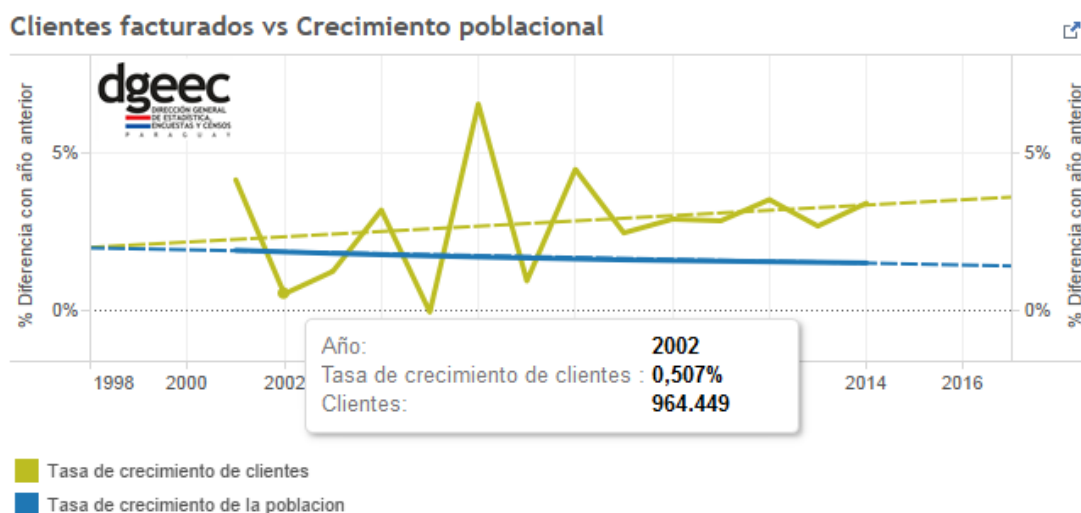


Figura 3.12: Clientes Facturados vs Crecimiento Poblacional

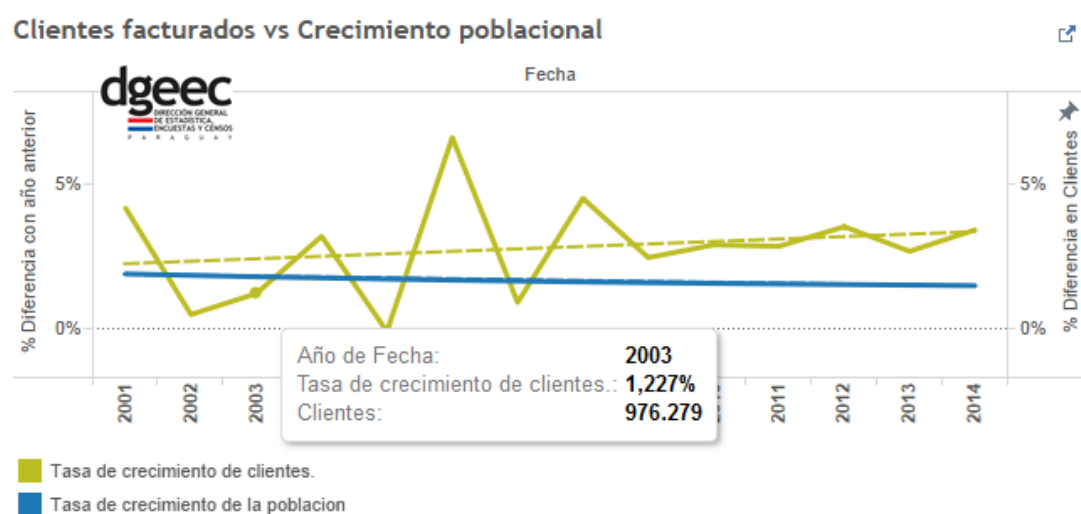


Figura 3.13: Clientes Facturados vs Crecimiento Poblacional

En el siguiente gráfico (ver Figura 3.14) se realiza una proyección o forecasting donde se muestra que existen altas probabilidades que si la entidad conserva la misma cantidad de clientes o estos crecen mínimamente, de igual manera el consumo podría aumentar o disminuir drásticamente. El aumento del consumo, podría deberse al aumento de productos electrónicos que consumen mucha más energía eléctrica y también a que el poder adquisitivo de cada ciudadano ha aumentado.

Este comportamiento se observa en la zona de cálculo de proyección (rojo, azul y amarillo suavizado).

Proyección de Clientes facturados, crecimiento poblacional y consumo de energía para los próximos 5 años.

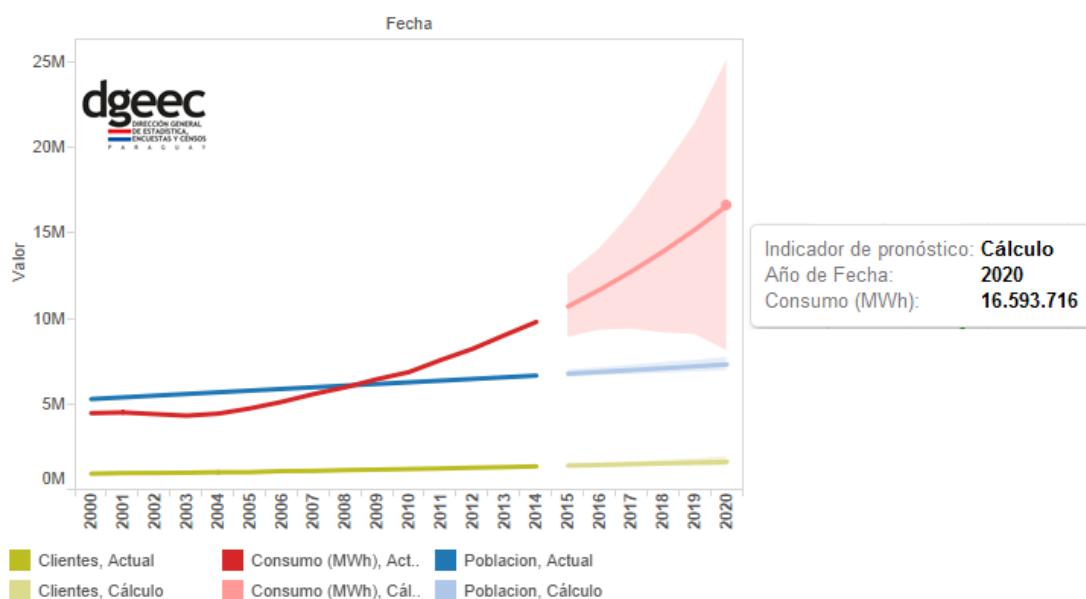


Figura 3.14: Proyección de Clientes Facturados Crecimiento Poblacional Y Consumo De Energía para los Próximos 5 Años

En el tercer y último gráfico de este panel, se muestra la porcentaje del crecimiento anual de los importes facturados y consumo de energía. Se puede observar que la facturación de la ANDE acompaña al consumo de energía eléctrica, exceptuando el año 2011(ver figura 3.15), en la cual el importe aumentó mas de lo que aumentó el consumo de energía, sin embargo en el año 2013(ver figura 3.16) el importe volvió a aumentar menos que antes.

Energía consumida vs Importe facturado

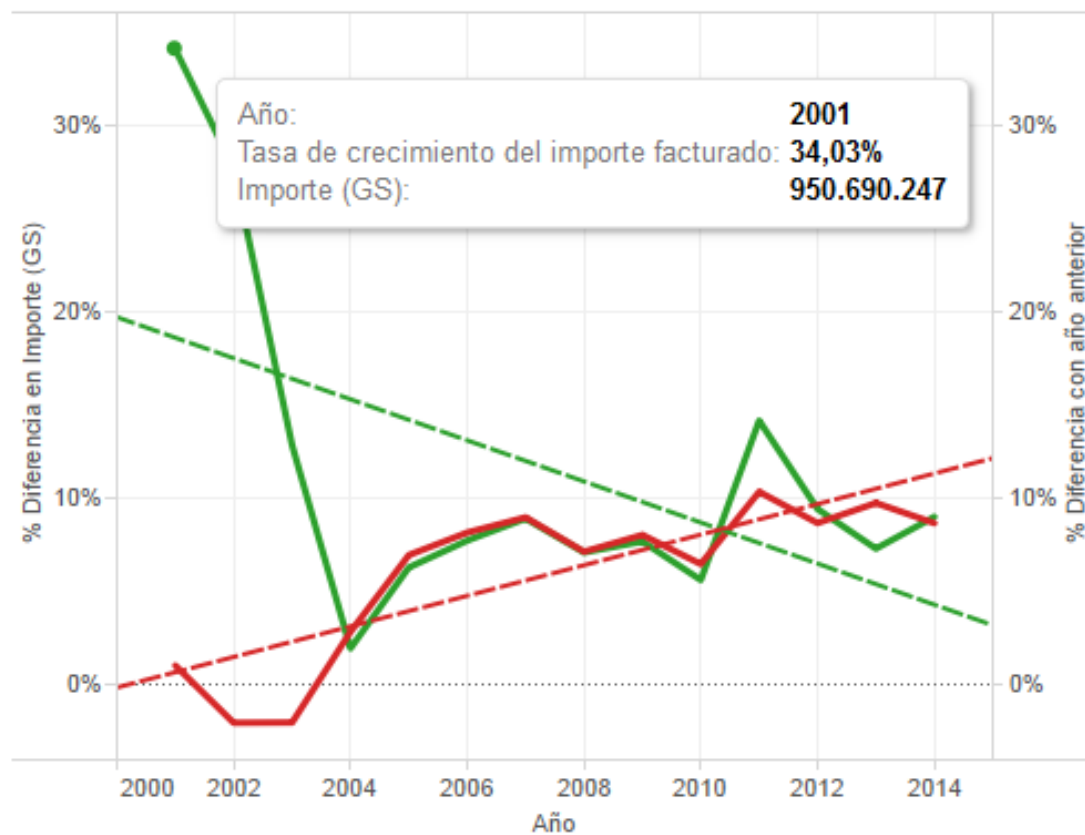


Figura 3.15: Energía Consumida vs Importe Facturado

Energía consumida vs Importe facturado

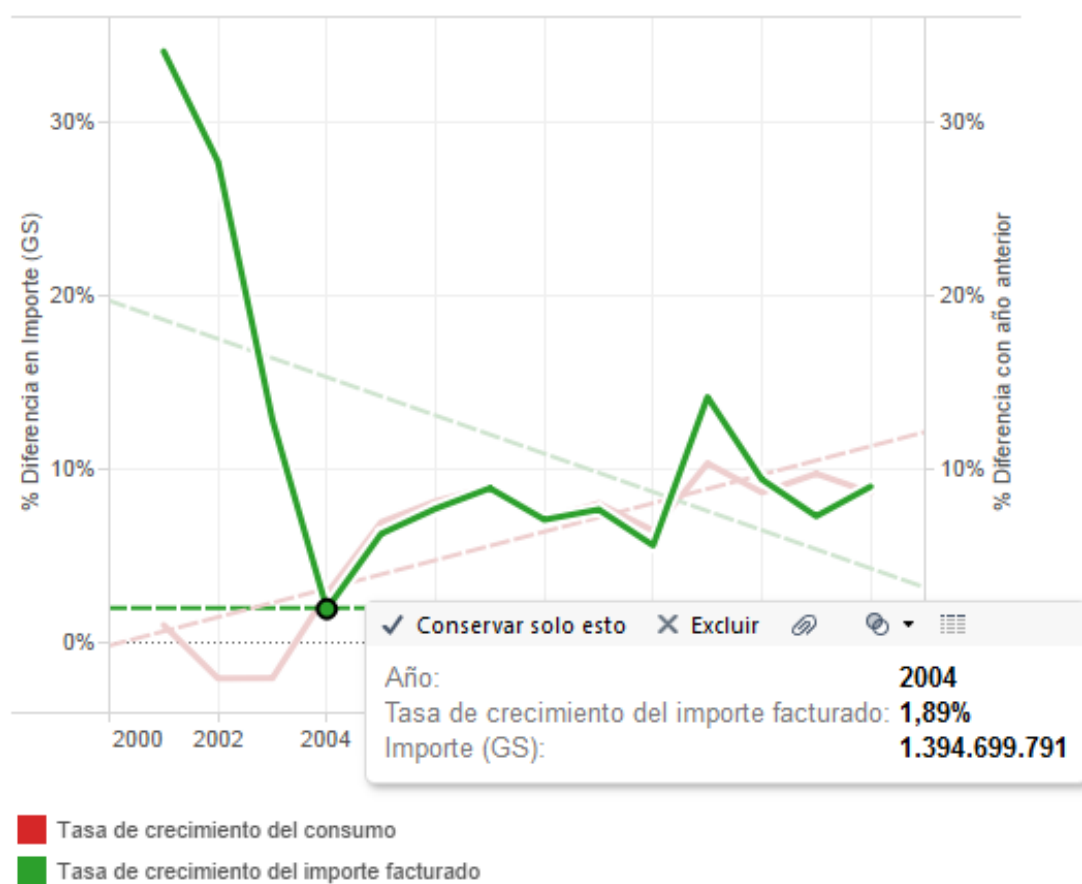


Figura 3.16: Energía Consumida vs Importe Facturado

3.2.2. Dashboard - Estadística de consumo de electricidad por sector

En el primer tab, “Consumo“, tenemos una estadística de consumo por grupo de consumidores de energía. Al abrir este panel, observamos que la el grupo de consumidores residencial es la que más demanda energía, históricamente. Otro dato interesante es la exportación de energía eléctrica, que según el gráfico, cada vez fue disminuyendo más. A nivel nacional la zona residencial se ubica en primer lugar con un 39,87 %. Luego viene el sector industrial, que consume el 22.11 % de la energía. El sector comercial es dueña del 17,39 %.



Figura 3.17: Estadística de consumo de electricidad por sector(1990-2014)

En el segundo tab, titulado “Facturación“, tenemos datos de facturaciones por grupo de consumidores. Al igual que el gráfico anterior observamos que la zona residencial es a la que más facturas se emiten.



Figura 3.18: Importe facturado por año y sector(1990-2014)

3.2.3. Dashboard - Tasa de crecimiento vs Consumo de energía

En este panel presentaremos informaciones sobre el consumo y clientes de la ANDE por grupo de consumidores y departamento.

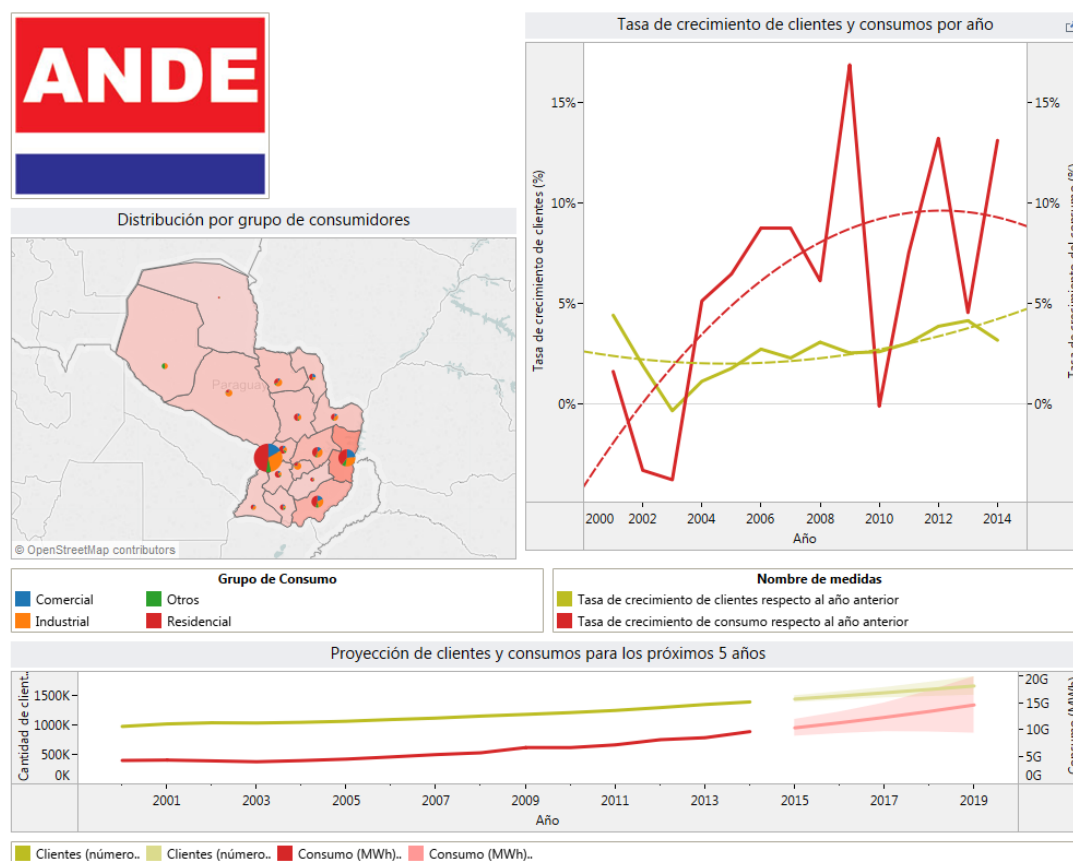


Figura 3.19: Tasa de crecimiento vs Consumo de energía

En el mapa se presentan gráficos de tortas por cada departamento, que representan los grupos de consumidores (Comercial, Industrial, Residencial y otros), donde cada color representa a un grupo en específico. Se puede ver que en la mayoría de los departamentos el grupo de consumo “residencial”, “industrial” y “comercial” son las que ocupan la mayor porción. Estos grupos sirven como filtro para los demás gráficos, por ejemplo, si se presiona sobre cualquier grupo de consumidores, los demás gráficos se actualizarán en tiempo real. Al ubicar el mouse sobre alguna porción de la torta, se muestra el consumo equivalente del departamento correspondiente.

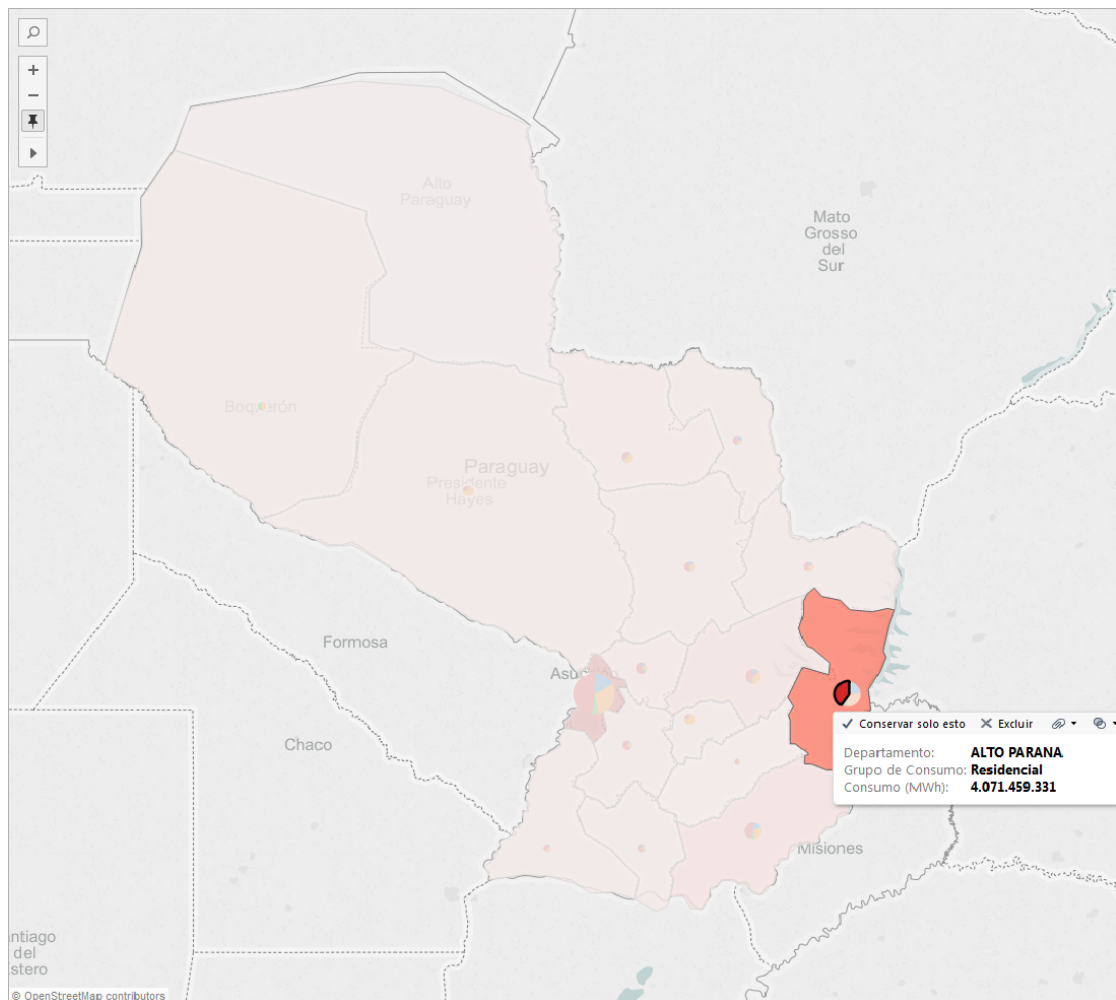


Figura 3.20: Tasa de crecimiento vs Consumo de energía, filtrado por el departamento Alto Paraná

En este gráfico se utilizó datos de los clientes de la entidad y el consumo de energía eléctrica. Se calcula la tasa de crecimiento anual de ambos. Observamos que el aumento de los clientes de la entidad es similar cada año, sin embargo, la tasa de crecimiento del consumo de energía es muy inestable. Existen ocasiones en que aumenta el 10 % o 20 % más cada año y también en la que disminuye la misma cantidad y esto sin que haya mucha variación en la cantidad de clientes.

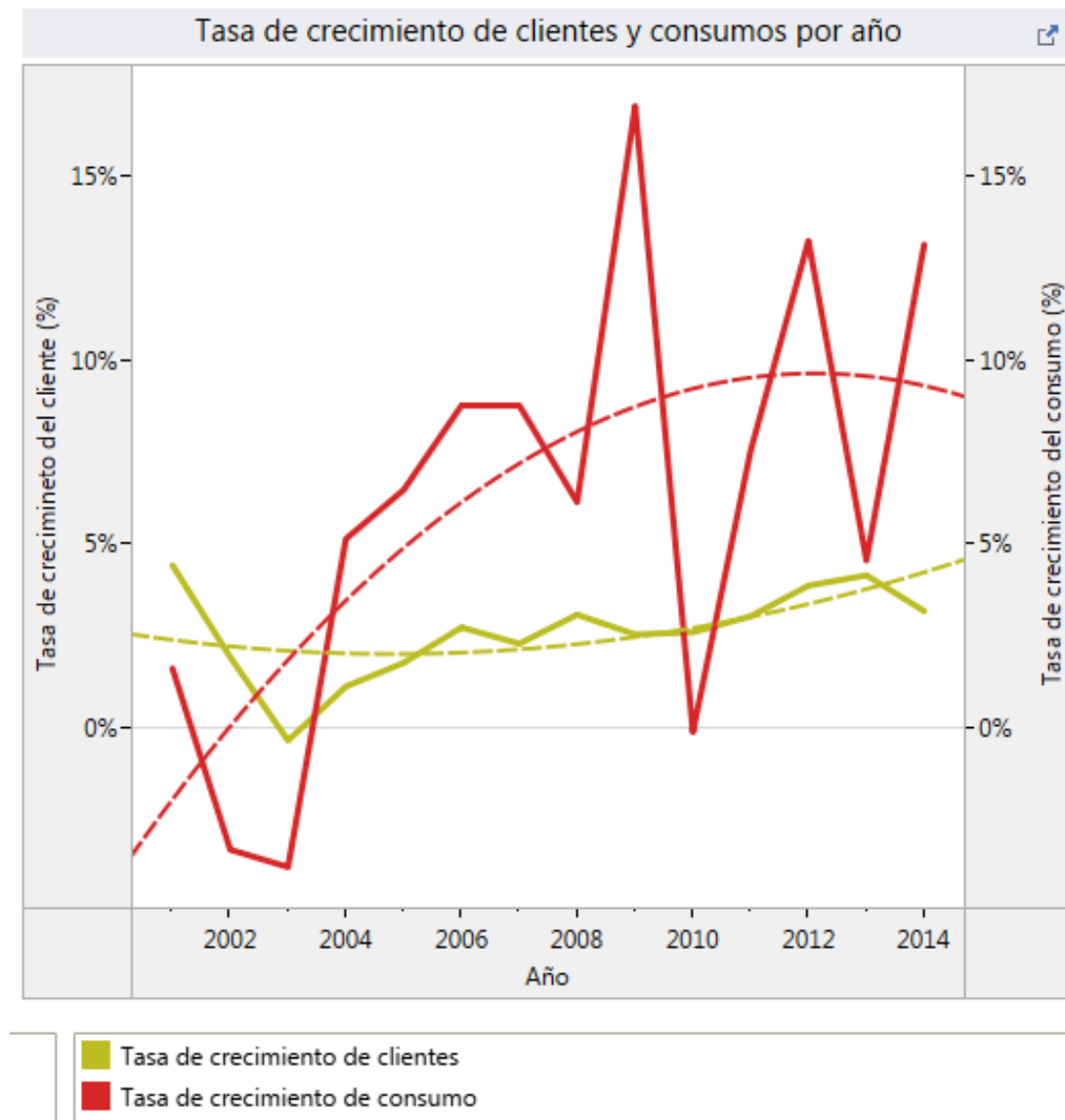


Figura 3.21: Tasa de crecimiento vs Consumo de energía

En este gráfico, titulado “Proyección de clientes y consumo para los próximos 5 años” se realiza una comparación entre los clientes y el consumo de energía eléctrica, similar que el gráfico anterior, pero con la diferencia de que en este gráfico se muestra la información con una unidad de medida diferente. A la izquierda tenemos el valor “Cantidad de clientes” y a la derecha el “consumo” en GWh.

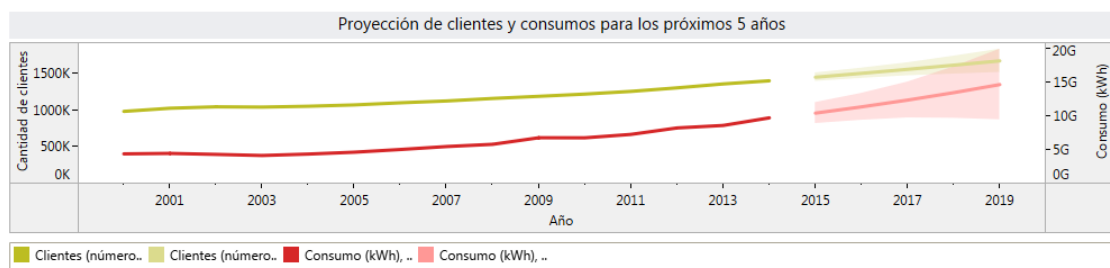


Figura 3.22: Proyección de clientes y consumos para los próximos 5 años

3.2.4. Dashboard - Tasa de crecimiento poblacional vs Consumo de energía año tras año

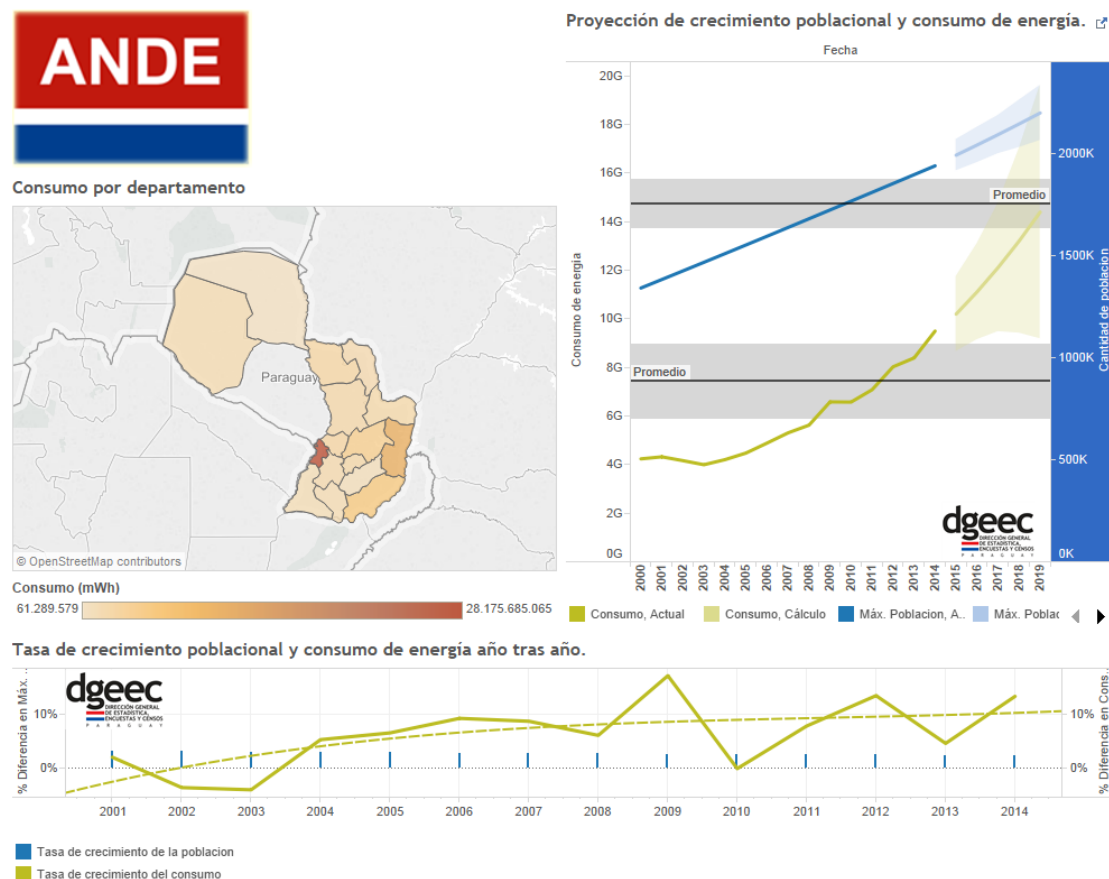
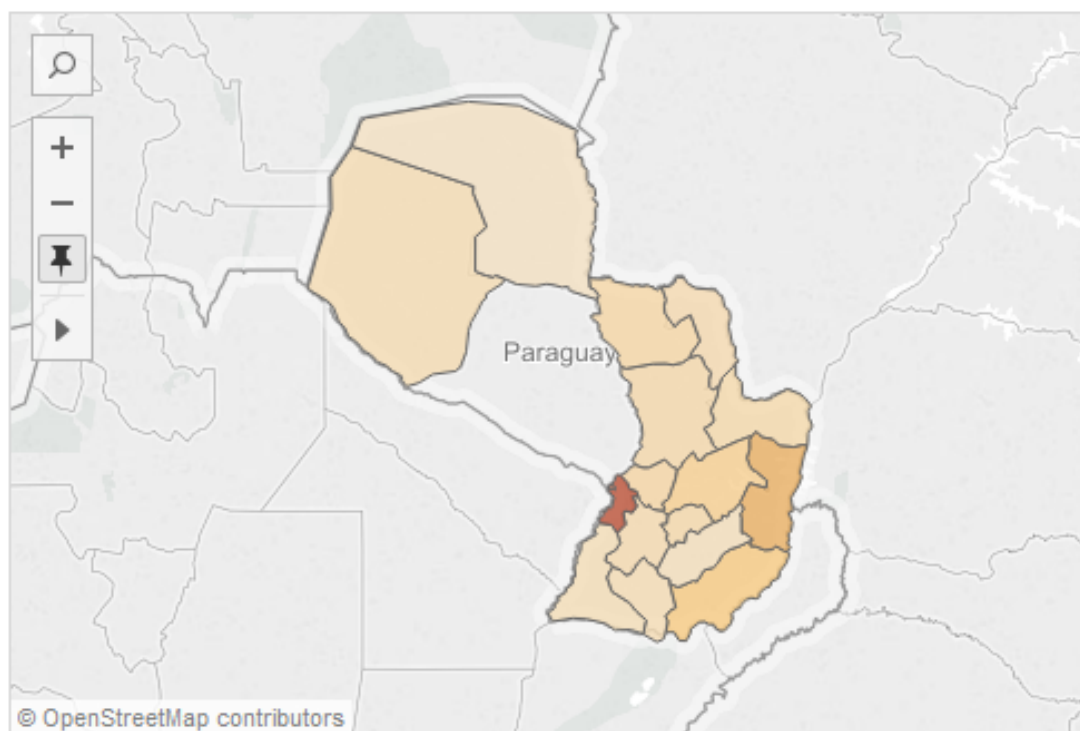


Figura 3.23: Dashboard - Tasa de crecimiento poblacional vs consumo de energía año tras año

En el mapa, donde el color más oscuro representa al departamento que consume más energía eléctrica y el color más claro, al que consume menos, vemos que los departamentos central, Alto Paraná son los que más demandan energía. Este tipo de gráfico es muy útil cuando la información se quiere analizar de forma macro y georeferenciada. Al ubicar el mouse sobre cualquier departamento, se muestra un pop up indicando el valor de consumo del departamento seleccionado. Al dar clic sobre un departamento los demás gráficos también se actualizarán en base a la selección.

Consumo por departamento



Consumo (kWh)

61.289.579



28.175.685.065

Figura 3.24: Consumo por departamento

En el segundo gráfico, titulado “Proyección de crecimiento poblacional y consumo de energía”, vemos el crecimiento de la población (números) y el crecimiento del consumo de energía eléctrica expresado en GWh. Al seleccionar un departamento en el mapa, se puede analizar esta información por cada uno de ellos.

Proyección de crecimiento poblacional y consumo de energía.

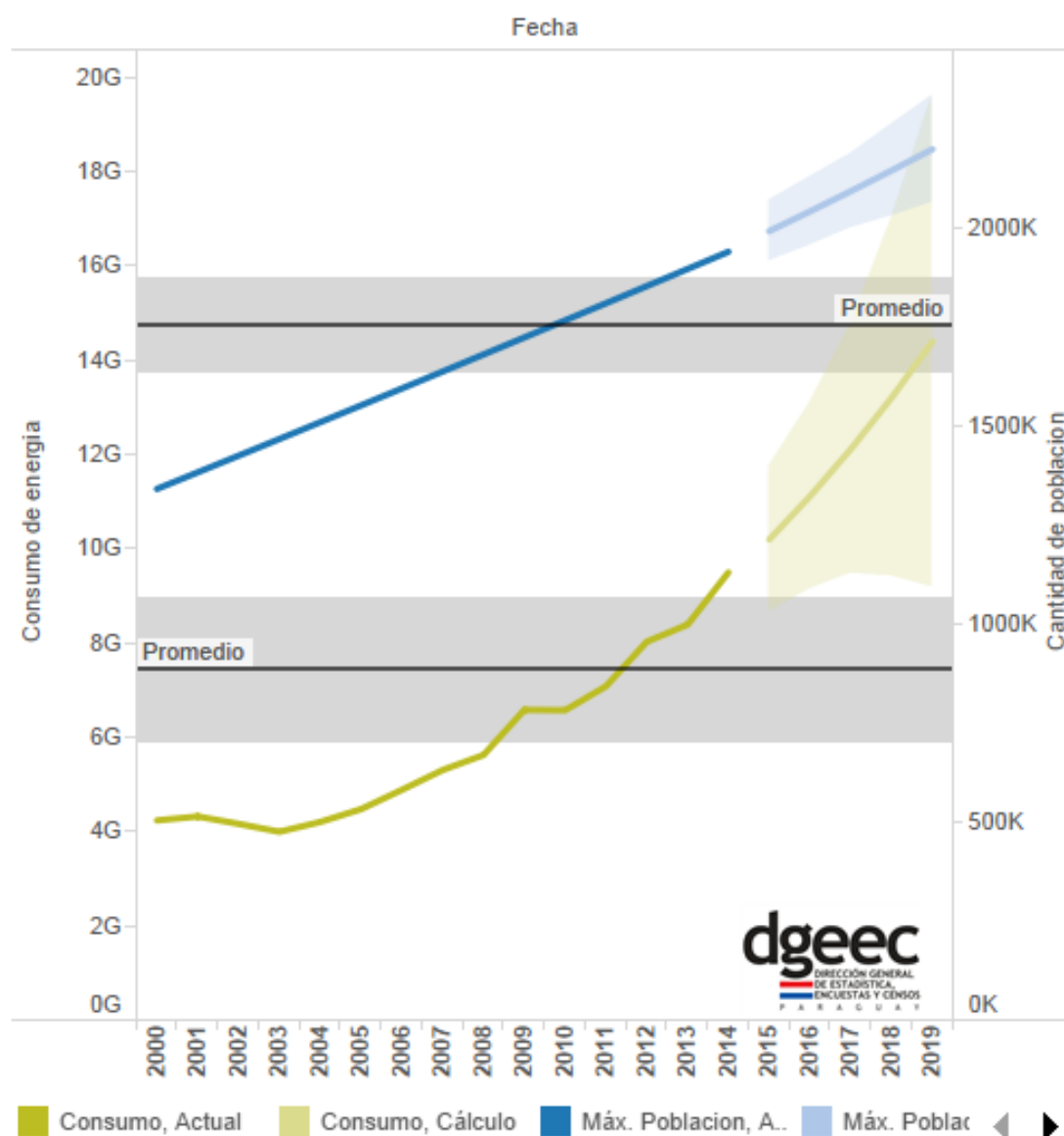


Figura 3.25: Proyección de crecimiento poblacional y consumo de energía

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.

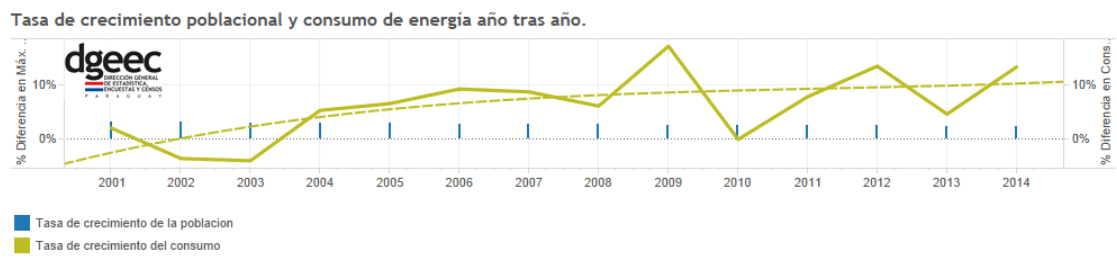


Figura 3.26: Tasa de crecimiento poblacional y consumo de energía años tras años

CAPITULO 4

4.1. Marco Metodológico

4.1.1. Alcance

En este trabajo 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.

4.1.2. 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).(Gómez, 2006)

4.1.3. 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.

CAPITULO 5

5.1. Conclusiones y Trabajos futuros

Referencias

- Adelman, S., Moss, L., y Barbusinski, L. (2002). I found several definitions of bi. *DM Review*, 5700–1.
- Business intelligence — Wikipedia, the free encyclopedia. (2015). *Wikipedia*. Descargado de https://en.wikipedia.org/wiki/Business_intelligence ([WEB; visitada 20-Octubre-2015])
- Data discovery — Wikipedia, the free encyclopedia. (2015). *Wikipedia*. Descargado de https://en.wikipedia.org/wiki/Data_discovery ([WEB; visitada 5-Octubre-2015])
- 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*, 14, 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), 20–33.
- Gómez, M. M. (2006). *Introducción a la metodología de la investigación científica*. Editorial Brujas.
- 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.
- 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.
- 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).