

UNIVERSIDAD CATO´ LICA “NUESTRA SEN˜ ORA DE LA ASUNCIO´ N” CAMPUS ALTO PARANA´

FACULTAD DE CIENCIAS Y TECNOLOG´IA

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

Iv´an Ariel C´aceres Can˜ete

Ariel Hern´an Landaida Duarte

Hernandar´ıas, abril de 2016

ii

UNIVERSIDAD CATO´ LICA “NUESTRA SEN˜ ORA DE LA ASUNCIO´ N” CAMPUS ALTO PARANA´

FACULTAD DE CIENCIAS Y TECNOLOG´IA

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

Iv´an Ariel C´aceres Can˜ete

Ariel Hern´an Landaida Duarte

Ricardo Luis Brunelli Montero, Ing.

Tutor

Hernandar´ıas, abril de 2016

iv

Iv´an Ariel C´aceres Can˜ete

Ariel Hern´an Landaida Duarte

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

Proyecto de Fin de Carrera presentado como requisito parcial para optar al t´ıtulo de Licenciado en Ana´lisis de Sistemas.

Facultad de Ciencias y Tecnolog´ıa, Universidad Cato´lica “Nuestra

Sen˜ora de la Asuncio´n”

Tutor: Ing. Ricardo Luis Brunelli Montero

Hernandar´ıas, abril de 2016

vi

Landaida Duarte, Ariel Herna´n;C

ceres Can˜ete, Iv´an Ariel.

(2016); DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY, una aplicacio´n para ayudar a las personas a analizar, visualizar y compartir informacio´n ra´pidamente. Hernandarias, Universidad Cato´lica. 110 p.

Tutor: Ing. Ricardo Luis Brunelli Montero.

Defensa de Proyecto de Fin de Carrera.

Palabras clave: Data Discovery, Business Intelligence.

Iv´an Ariel C´aceres Can˜ete

Ariel Hern´an Landaida Duarte

DATA DISCOVERY APLICADOS A DATOS DEL PARAGUAY

Proyecto de Fin de Carrera presentado como requisito parcial para optar al t´ıtulo de Licenciado en Ana´lisis 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:

Hernandar´ıas, abril de 2016

viii

DEDICATORIA

A mis padres, por la oportunidad de existir, por su sacrificio en algu´n tiempo incomprendido, por su ejemplo de superacio´n incansable, por su comprensio´n y confianza, por su amor y amistad incondicional, por los consejos que siempre me han dado para sobrellevar los desaf´ıos 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, dedicacio´n, motivacio´n, criterio y aliento. Ha hecho facil lo dif´ıcil. Ha sido un privilegio poder contar con su gu´ıa y ayuda.

A nuestra familia quienes por ellos somos lo que somos. A nuestros padres por su apoyo, consejos, comprensio´n, amor, ayuda en los momentos dif´ıciles, y por ayudarnos con los recursos necesarios para estudiar. Nos han dado todo lo que somos como persona, valores, principios, cara´cter, empen˜o, perseverancia, y coraje para conseguir nuestros objetivos.

Iv´an Ariel C´aceres Can˜ete

x

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, dedicacio´n, motivacio´n, criterio y aliento. Ha hecho facil lo dif´ıcil. Ha sido un privilegio poder contar con su gu´ıa y ayuda.

A nuestra familia quienes por ellos somos lo que somos. A nuestros padres por su apoyo, consejos, comprensio´n, amor, ayuda en los momentos dif´ıciles, y por ayudarnos con los recursos necesarios para estudiar. Nos han dado todo lo que somos como persona, valores, principios, cara´cter, empen˜o, perseverancia, y coraje para conseguir nuestros objetivos.

Ariel Hern´an Landaida Duarte

xii

RESUMEN

Hasta hace poco tiempo, la mayor´ıa de las organizaciones prove´ıan datos es- tructurados, limpios, e integrados, resumidos a niveles convenientes para plata- formas 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 volu´menes de datos directamente de sus sitios y aplicaciones web, hoy generalmente denomina- do Big Data. Aquellos datos obtenidos, en su mayor´ıa, eran heterog´eneos y hasta inclusive no estructurados, y esa situaci´on genero´ la necesidad de crear otro tipo de herramienta que ayude al tomador de decisi´on en la bu´squeda de patrones y relaciones. Este nuevo enfoque, denominado Data Discovery, no pod´ıa ser igual a las t´ecnicas ya tradicionales, adema´s deb´ıa tener caracter´ısticas como innovaci´on 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 espec´ıficamente Data Discovery. Se aplican estas t´ecnicas a datos de dos instituciones del estado, demostrando los beneficios de aplicar este tipo de t´ecnica.

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 gene- rally 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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2.1. Etapas de BI como fuentes de datos de calidad | |  | . | . | . | . | 5 |
| 2.2. Framework Analitico de Negocios Gartner . | | . | . | . | . | . | 16 |
| 2.3. Espectro Analitico . . . . . . . . . . | | . | . | . | . | . | 17 |
| 2.4. Tipico uso de estilos an´aliticos . . . . . . | | . | . | . | . | . | 19 |
| 2.5. | Caracter´ısticas de BI en niveles y Plataforma Anal´ıtica | | | |  | . | 23 |
| 3.1. | Cuadrante M´agico para BI y Plataformas Anal´ıticas | | | | . | . | 25 |
| 3.2.  3.3. | Puntuaciones de producto o servicio para ana´lisis descen-  tralizado . . . . . . . . . . . . . . . . . . Puntuaciones de producto o servicio guiados por Data Dis- covery . . . . . . . . . . . . . . . . . . . . | | | | | | 26  27 |
| 3.4. | Arrastre el campo pa´ıs para el campo desplegable sen˜alado. | | | | | | 28 |
| 3.5. | Arrastrar hojas de trabajo al dashboard. . . . | | . | . | . | . | 29 |
| 3.6. | Clientes Facturados vs Crecimiento Poblacional | | . | . | . | . | 31 |
| 3.7. | Clientes Facturados vs Crecimiento Poblacional | | . | . | . | . | 32 |
| 3.8. | Clientes Facturados vs Crecimiento Poblacional | | . | . | . | . | 32 |
| 3.9. | Clientes Facturados vs Crecimiento Poblacional | | . | . | . | . | 33 |
| 3.10. | Clientes Facturados vs Crecimiento Poblacional | | . | . | . | . | 34 |
| 3.11. | Clientes Facturados vs Crecimiento Poblacional | | . | . | . | . | 34 |
| 3.12. | Proyecci´on de Clientes Facturados Crecimiento Poblaciona  Y Consumo De Energ´ıa para los Pr´oximos 5 An˜os . . | | | | | l  . | 35 |
| 3.13. | Energ´ıa Consumida vs Importe Facturado . . . . . | | | | | . | 36 |
| 3.14. | Energ´ıa Consumida vs Importe Facturado . . . . . | | | | | . | 37 |

3.15. Estad´ıstica de consumo de electricidad por sector(1990-2014) 38

3.16. Importe facturado por an˜o y sector(1990-2014) . . . . 39

3.17. Tasa de crecimiento vs Consumo de energ´ıa . . . . . . 40

3.18. Tasa de crecimiento vs Consumo de energ´ıa, filtrado por el

departamento Alto Parana´

. . . . . . . . . . . . 41

3.19. Proyecci´on de clientes y consumos para los pro´ximos 5 an˜os 42

3.20. Panel de Tasa de crecimiento poblacional y Consumo de energ´ıa anual . . . . . . . . . . . . . . . . . 42

3.21. Consumo por departamento . . . . . . . . . . . . 43

3.22. Proyecci´on de crecimiento poblacional y consumo ee energ´ıa 44

xvii

LISTA DE TABLAS

2.1. T´ecnicas actuales de BI . . . . . . . . . . . . . . . . 12

´INDICE

1. CAPITULO 1 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.1. | Introduccio´n . . . . . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 1 |
| 1.2. | Planteamiento del problema | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 1 |
| 1.3. | Objetivos . . . . . . . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 2 |
|  | 1.3.1. Objetivo General . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 2 |
|  | 1.3.2. Objetivos Espec´ıficos |  | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 2 |
| 1.4. | Justificaci´on . . . . . . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 3 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2. | CAPITULO 2 |  |  |  |  |  | 3 |
|  | 2.1. MARCO TEO´ RICO . . . . . . . . . . . . | . | . | . | . | . | 4 |
|  | 2.1.1. Business Intelligence . . . . . . . . . . | . | . | . | . | . | 4 |
|  | 2.1.2. Data Discovery Analysis . . . . . . . . | . | . | . | . | . | 13 |
|  | 2.1.3. Framework de Ana´lisis de Negocio de Gartner | . | . | . | . | . | 15 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3. | CAPITULO 3 |  |  |  |  |  |  |  | 24 |
|  | 3.1. Seleccio´n de la herramienta para Data Discovery | . | . | . | . | . | . | . | 24 |
|  | 3.1.1. Cuadrante Ma´gico de Gartner . . . . | . | . | . | . | . | . | . | 24 |

3.2. Aplicaci´on de Data Discovery a datos de instituciones del Estado . 29

3.2.1. Datos de la ANDE y de la DGEEC . . . . . . . . . 29

3.2.2. Dashboard de control / monitoramiento . . . . . . . 30

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 4. | CAPITULO 4 |  |  |  |  |  | 46 |
|  | 4.1. Marco Metodolo´gico . . . . . . . . . . . . . | . | . | . | . | . | 46 |
|  | 4.1.1. Alcance . . . . . . . . . . . . . . . | . | . | . | . | . | 46 |
|  | 4.1.2. Enfoque . . . . . . . . . . . . . . | . | . | . | . | . | 46 |
|  | 4.1.3. T´ecnica e Instrumentos de recolecci´on de datos | . | . | . | . | . | 46 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5. | CAPITULO 5 |  |  |  |  |  |  | 46 |
|  | 5.1. Conclusiones y Trabajos futuros | . . . . . . . . | . | . | . | . | . | 47 |
|  | Referencias . . . . . . . . . . | . . . . . . . . | . | . | . | . | . | 50 |

xix

xx

CAPITULO 1

1.1. Introduccion

En este trabajo fue elaborada una propuesta de estado del arte del

a´rea de

BI (Business intelligence) y espec´ıficamente Data Discovery. El objetivo fue apli- car los conocimientos adquiridos al entorno nacional, en este caso aplicados a una institucio´n del estado (Administracio´n Nacional de Electricidad, ANDE), para el apoyo en la toma de decisiones. Aplicaremos las t´ecnicas de Data Discovery a los datos de dos instituciones del estado, espec´ıficamente la Administracion Nacional de Electricidad (ANDE) y Direccion General de Estad´ısticas, Encuestas y Cen- so (DGEEC), donde demostraremos que con datos de calidad podr´ıamos detectar oportunidades que nos faciliten la toma de decisiones en la instituci´on. Utiliza- remos conjuntos de datos de las instituciones mencionadas m´as arriba para este fin. Este trabajo est´a estructurado de la siguiente forma: El cap´ıtulo I presenta el planteamiento del problema, los objetivos generales y espec´ıficos, concluyendo con la justificaci´on. El cap´ıtulo II presenta el marco teorico. Se compone de con- ceptos, componentes, tecnolog´ıas de BI, y el framework de Gartner. El cap´ıtulo III presenta los diferentes paneles indicadores (dashboards) realizados con la he- rramienta seleccionada para aplicar data discovery sobre los conjuntos de datos existentes. Se explica el cuadrante m´agico de Gartner, el cual apoyo´ a la seleccio´n de la herramienta mencionada. Adema´s se presentan ejemplos de ana´lisis de da- tos y la informacio´n obtenida, como resultado de la combinaci´on del uso de las herramientas y datos.

1.2. Planteamiento del problema

Actualmente la mayor´ıa de las instituciones pu´blicas e incluso las del sector pri- vado, tienen un bajo nivel de inversio´n en tecnolog´ıa. Generalmente en empresas que exigen toma de decisi´on, con frecuencia optan por decisiones de negocios no o´ptimas debido a que no poseen la suficiente experiencia o suficiente datos procesa- dos del negocio para llegar al correcto ana´lisis, o pueden estar usando herramientas incorrectas. Teniendo u´nicamente la experiencia como herramienta, puede ser su- ficiente s´olo en organ

izaciones pequen˜as, donde se adquiere conocimiento sin necesidad de alguna herramienta de ana´lisis de datos, por ejemplo, sabriamos cuales son los productos ma´s vendidos o m´as rentables en un negocio pequen˜o, y esto se complica a medida que la organizaci´on 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 infor- maci´on. Estas informaciones son obtenidas a trav´es de un ana´lisis profundo de datos, por consiguiente estos datos tambi´en deben ser de buena calidad. Si una organizaci´on posee la capacidad de obtener lo que desea por medio de los datos, es

seguro que su crecimiento ser

positivo, debido a que tomara´ mejores decisiones,

y estos ayudar´ıan a su evoluci´on y estabilidad a lo largo del tiempo.

Existen ocasiones en que la organizacio´n posee suficientes datos pero no con- sigue analizarlos o procesarlos por falta de conocimiento del negocio, adema´s de desconocimiento t´ecnico de las herramientas.

1.3. Ob jetivos

1.3.1. Ob jetivo General

En este trabajo se aplican las t´ecnicas de Data Discovery a datos de dos ins- tituciones del estado: ANDE y DGEEC. Se proponen algunos tableros de control (Dashboards) con indicadores, para demostrar que con los datos prove´ıdos es po- sible detectar situaciones, y estimar o visualizar acontecimientos de inter´es para una organizacio´n.

1.3.2. Ob jetivos Espec´ıficos

Estudio del estado del arte de BI y Data discovery.

Selecci´on de la herramienta adecuada para este trabajo.

Ana´lisis y cruzamiento de datos para el descubrimiento de situaciones de inter´es.

Elaboracio´n de dashboards que reflejen el an´alisis de los datos con la herra- mienta seleccionada.

Elaboracio´n de gr la instituci´on.

ficos de tendencias, proyecciones de consumo y clientes de

1.4. Justificacion

Falta de un ambiente anal´ıtico corporativo, que proporcione informaciones e indicadores necesarios para la toma de decisi´on. Falta de agilidad para elaborar informes con indicadores de tendencias de consumo de energ´ıa, que apoye a la planificaci´on territorial de expansio´n de la red de transmisi´on el´ectrica. Falta de

indicadores de consumo geogr

fico de electricidad, relacionados con indicadores po-

blacionales, para apoyo en la planificaci´on de inversio´n en el aumento de capacidad de transformadores por regi´on.

CAPITULO 2

2.1. MARCO TEO´ RICO

2.1.1. Business Intelligence

Business Intelligence (BI) o Inteligencia de negocio, es definido como la capa- cidad de una organizacio´n para tomar todos sus procesos y capacidades, y luego convertirlos en conocimiento, o en otras palabras, obtener informacio´n correcta pa- ra la persona correcta, en el tiempo correcto, a trav´es del canal correcto (Kumari,

2013).

Esto produce grandes cantidades de informacio´n que pueden llevar al desarrollo de nuevas oportunidades para la organizacio´n. La identificacio´n e implementaci´on de estas oportunidades en un estrategia efectiva, puede proporcionar una ventaja competitiva de mercado y la deseada estabilidad a largo plazo (Rud, 2009).

Las tecnolog´ıas de BI ofrecen vistas hist

ricas, actuales y predictivas de las ope-

raciones de una empresa. Algunas de las funcionalidades ma´s comunes de BI son reportes, procesamiento anal´ıtico en l´ınea (en adelante OLAP, Analytical Proces- sing), miner´ıa de datos, miner´ıa de procesos, procesamiento de eventos complejos, gesti´on del rendimiento empresarial, benchmarking, miner´ıa de texto, ana´lisis pre- dictivos y ana´lisis prescriptivos. BI tiene como objetivo apoyar una mejor toma de decisiones, por tanto, un sistema de BI puede tambi´en ser llamado sistema de apoyo a las decisiones o DSS (del ingl´es, decision support system)(Rud, 2009).

Las herramientas de BI son disen˜adas para obtener, analizar y reportar datos. Estas herramientas generalmente leen datos que han sido previamente almace- nados, frecuentemente, aunque no necesariamente en un almac´en 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, m´etodos, y tecnolog´ıas para convertir datos separados o aislados de una organizacio´n en informacio´n u´til 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 trans- versales a toda la organizacio´n. El proceso de BI y sus varias etapas son mostrados en la Figura 2.1. La calidad de los datos juegan un rol cr´ıtico e importante en el

´exito 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 organizaci´on, adem´as de impedir el crecimiento de la organizaci´on (Kumari, 2013).

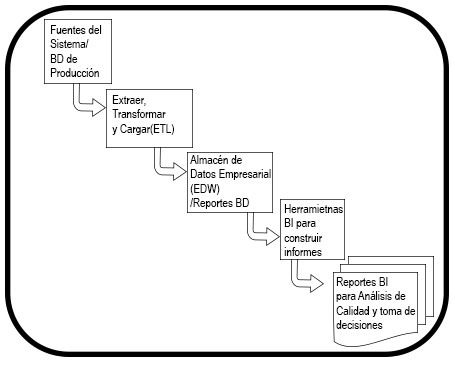


Figura 2.1: Etapas de BI como fuentes de datos de calidad

Una cuesti´on fundamental es el hecho que una organizaci´on corre sobre datos; y actu´a como insumo para el motor de la industria corporativa. Una organizaci´on no puede comprender a sus clientes, proveedores, competidores o a su propia gen- te, procesos, y rendimiento sin datos de buena calidad. Por consiguiente, la alta direccio´n de una empresa y el ´area de TI (tecnolog´ıa de la informaci´on) deber´ıan 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 organizacio´n pueden explorar y realizar cortes de datos, usando herramientas sofisticadas que permiten la navegaci´on de dimensiones tales como el tiempo o jerarqu´ıas. OLAP, provee vistas resumidas multidimensionales de datos del negocio de una organizacio´n, y es usado para reportes, an´alisis, modelado y planificacio´n para la optimizacio´n de una organizaci´on.(Malhotra, 2001)

Las t´ecnicas y herramientas OLAP pueden ser usados para trabajar con da-

tawarehouse o con data marts (un subconjunto de datos de un

a´rea espec´ıfica)

disen˜ados para sistemas sofisticados. Este tipo de consultas son requeridas para descubrir tendencias y analizar factores cr´ıticos. Los reportes generan vistas agre- gadas de datos para mantener la gestio´n informada sobre el estado de sus negocios. Otras herramientas de BI son usadas para almacenar y analizar datos, tales como la miner´ıa de datos y el datawarehouse; sistemas de soporte o apoyo a las deci- siones y previsiones; almac´en de documentos y gesti´on de documentos; gesti´on del conocimiento; mapeamiento, visualizacio´n de informacio´n y paneles (dashboards); sistemas de informaci´on de gestio´n, sistemas de informacio´n geogr´aficas; ana´lisis de tendencias; Software como servicio (SaaS) y otros.(Malhotra, 2001)

2.1.1.1.2. An´alisis Avanzado Ana´lisis Avanzado: conocido como miner´ıa de datos y an´alisis predictivos, toma las ventajas de las t´ecnicas de an´alisis estad´ıs- ticos para predecir o proveer medidas de certeza sobre ciertos hechos. La gesti´on sobre el rendimiento de una organizacio´n (Portales, cuadros de mando, paneles de control): esta categor´ıa general normalmente provee un sistema de varios com- ponentes interconectados, de tal forma que en conjunto describan una historia. Por ejemplo, un cuadro de mando integral que muestre componentes de indica- dores financieros, todos ellos combinados pueden describir m´etricas y patrones de aprendizaje y crecimiento en las organizaciones.(Gangadharan y Swami, 2004)

2.1.1.1.3. BI en tiempo real: permite la distribuci´on de m´etricas en tiempo real a trav´es de emails, sistemas de mensajer´ıa instant´anea y/o pantallas interac- tivas.

2.1.1.1.4. Datawarehouse y Datamarts Datawarehouse y Datamarts: El datawarehouse es un componente importante de BI. El datawarehouse soporta la propagacio´n f´ısica de los datos manejando grandes volu´menes de registros de las or- ganizaciones para integraci´on, limpieza, agregacio´n y tareas de consulta.(Gangadharan y Swami, 2004)

Tambi´en puede contener datos operacionales, los cuales pueden ser definidos como un conjunto actualizable de datos integrados disponibles para toda una or- ganizacio´n, para la toma de decisiones t´acticas de un asunto espec´ıfico. Contiene datos vivos actualizados en tiempo real, no solamente fotos de un momento espe-

c´ıfico, y tambi´en conserva un historial m´ınimo. Las fuentes de datos pueden ser bases de datos operacionales, datos histo´ricos, datos externos, por ejemplo, de or- ganizaciones de investigaci´on de mercados o desde internet mismo, o informacio´n 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 organizaci´on. (Gangadharan y Swami,

2004)

Un datamart, tal como se describe en (Kumari, 2013), es una colecci´on 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 as´ı sucesivamente.

Cada departamento tiene su propia interpretacio´n de c´omo debe verse un data mart y el data mart de cada departamento es particular y atiende las necesida- des espec´ıficas del a´rea. Similar al datawarehouse, los data marts contienen datos transaccionales que ayudan a expertos en negocios a crear una estrategia basada en el ana´lisis de las tendencias y experiencias pasadas. La principal diferencia es que la creacio´n de los data marts se basa en una necesidad espec´ıfica, predefinida para un grupo determinado. Un data mart puede apoyar o soportar procesos o unidades de negocio espec´ıficos (Ranjan, 2009).

Las herramientas de BI son ampliamente aceptadas como una capa intermedia entre aplicaciones transaccionales y aplicaciones de apoyo a la toma de decisio´n,

´estas esta´n desacopladas y extraen informaciones de transacciones de negocio. Las habilidades de BI incluyen, apoyo a la decisio´n, procesamiento anal´ıtico en l´ınea, an´alisis estad´ısticos, ana´lisis predictivo, y la miner´ıa de datos (Ranjan, 2009).

Las fuentes de datos pueden ser bases de datos operacionales, datos hist

ricos,

datos externos por ejemplo, desde las empresas de investigaci´on de mercados o desde internet, datos no estructurados de redes sociales, o informacio´n desde un datawarehouse existente. Las fuentes de datos pueden ser bases de datos relaciona- les o cualquier otra estructura de datos de sistemas transaccionales. Ellos tambi´en pueden residir en muchas plataformas diferentes, tales como tablas, hojas de c´alcu- lo, o informaci´on no estructurada, tales como archivos de texto plano o im´agenes y otras informaciones multimedia (Ranjan, 2009).

2.1.1.2. Beneficios de BI

BI provee muchos beneficios a las compan˜´ıas que lo utilizan. Puede ayudar a eliminar muchas conjeturas err´oneas dentro de una organizacio´n, mejorando la co- municacio´n entre departamentos mientras se coordinan las actividades, y apoyan a las organizaciones para responder ra´pidamente a cambios de condiciones finan- cieras o preferencias de clientes. BI puede ayudar a mejorar el rendimiento general de una organizacio´n (Ranjan, 2009).

La informaci´on es frecuentemente considerada como el segundo recurso m´as

importante que una compan˜´ıa tiene (lo ma´s valorable de una compan˜´ıa son las personas). Cuando una compan˜´ıa puede tomar decisiones basadas en informaci´on oportuna y precisa, puede ayudar a mejorar su rendimiento en su segmento de mercado. BI tambi´en agiliza la toma de decisiones, ayuda a actuar r´apida y correc- tamente con la informacio´n adecuada antes de otras empresas de la competencia. Tambi´en pueden mejorar la experiencia del cliente, teniendo en cuenta la respuesta oportuna y adecuada a los problemas y prioridades de los mismos. A continuacio´n se listan algunos de estos beneficios:

Con herramientas de BI, los empleados pueden facilmente convertir sus cono- cimientos de negocio en inteligencia anal´ıtica para resolver muchas cuestiones de negocio, tales como incrementar la tasa de respuestas desde correos elec- tr´onicos, tel´efonos, y mejorar las campan˜as de ventas desde internet.

Con BI, las empresas pueden identificar sus clientes ma´s rentables y las razo- nes subyacentes para la lealtad de esos clientes, as´ı como identificar clientes futuros con grandes potenciales.

Analizar los datos de clics para mejorar las estrategias de comercio electr´o- nico.

Deteccio´n r

pidamente de problemas reportados de productos para minimi-

zar el impacto de las deficiencias en sus disen˜os. Descubrir lavado de dinero de actividades delictivas.

Analizar la rentabilidad potencial del cliente, y reducir el riesgo a trav´es de una puntuacio´n ma´s precisa de cr´edito financiero de los mismos.

Determinar cuales son las combinaciones de productos y servicios que los clientes son ma´s propensos a comprar y cuando.

Analizar los ensayos cl´ınicos de farmacos experimentales. Establecer tarifas m´as rentables para las primas de seguros.

Reducir el tiempo fuera de un equipamiento mediante la aplicacio´n de man- tenimiento predictivo.

Determinar con el ana´lisis de desercio´n y rotacio´n 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 cr´edito o tarjetas telefonicas son robadas. Identificar nuevos compuestos de farmacos moleculares prometedores.

2.1.1.3. Tecnolog´ıa de BI

La inteligencia empresarial provee datos organizacionales de tal manera que

los filtros de conocimientos organizacionales puedan f

cilmente asociarse con estos

datos y volverlos en informaci´on para la organizaci´on. Las personas involucradas en procesos de inteligencia de negocios podr´ıan usar software y otras tecnolog´ıas para reunir, almacenar, analizar, proveer accesos a datos, y presentar esos datos de una manera simple y u´til.

El software ayuda en la gestio´n de una organizacio´n, y a las personas a hacer mejores decisiones de negocios, teniendo la informacio´n precisa, actualizada, y re- levante cuando lo necesiten. Algunas empresas usan data warehouse porque es un conjunto de informaci´on lo´gica 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 t´ecnicas que deber´ıan ser tratadas:

Seguridad y acceso de usuarios al data warehouse. Volumen de datos (capacidad).

Cua´nto tiempo ser´a almacenado el dato (retencio´n de datos). Sizing y rendimiento de infraestructura (servidores).

Las personas que trabajan en BI desarrollan productos que facilitan el traba- jo, especialmente cuando las tareas de inteligencia involucran conseguir y analizar grandes cantidades de datos no estructurados. Cada proveedor t´ıpicamente define BI de una forma particular, y comercializa herramientas para hacer BI de la for- ma en que cada uno lo propone. BI incluye herramientas en diversas categor´ıas, incluyendo las siguientes: (Ranjan, 2009).

AQL (Associative Query Logic) - L´ogica Asociativa de Consultas. M´etricas y mediciones del rendimiento del negocio.

Planeamiento Empresarial.

Data mining (DM), Data Farming, y Data warehouses. Sistemas de apoyo a la decisi´on (DSS) y predicci´on. Datawarehouse de documentos y gestio´n documental. Sistema de Gesti´on Empresarial.

Finanzas y presupuestos. Recursos humanos. Gesti´on del conocimiento.

Mapeamiento, visualizacio´n de la informaci´on, y paneles de control (dash- boards).

Sistemas de gestio´n de informaciones. Sistemas de informaci´on geogr´afica (GIS).

OLAP (Online Analytical Processing) y an´alisis multidimensional; a veces simplemente llamado “Analytics“ (basado tambi´en en “hipercubo“ o “cubo“). BI en tiempo real.

Ana´lisis de datos estad´ısticos y t´ecnicos.

Gesti´on de la l´ınea de produccio´n, Gesti´on de demandas.

Gesti´on de la cadena de Suministro/Gestio´n de la cadena de demanda. Ana´lisis 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 acci´on. Mas y m´as organizaciones han comenzado a disponibilizar ma´s datos

con mayor velocidad. En el pasado, los datos s

lo estaban disponibles despu´es 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. Recientemen- te, los bancos han intentado disponibilizar los datos en el intervalo m´as corto y reduciendo los atrasos (Ranjan, 2009).

Por ejemplo, para negocios de alto riesgo operacional (por ejemplo, tarjetas de cr´editos), un banco multinacional disponibiliza los datos relacionados con KPI semanalmente, y en ocasiones ofrece un an´alisis diario de los nu´meros. Esto signi- fica que los datos normalmente esta´n disponibles a cada 24 horas, requiriendo la automatizacio´n y el uso de sistemas de TI.

2.1.1.4. Breve discucion

La experiencia actual de cualquier nueva forma de organizacio´n 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 cr´ıticas relacionada a la cadena de valor. Sin un BI eficaz para dirigir las organizaciones orientadas a los procesos de apoyo, esto no ser´ıa posible.

(Davenport, 1993) describe varias cuestiones en la reingenier´ıa en las innova- ciones de los procesos de negocio. De acuerdo a (Adelman, Moss, y Barbusinski,

2002), BI es un t´ermino que engloba un amplio rango de software de ana´lisis y soluciones para recolectar, consolidar, analizar y proveer acceso a la informaci´on de una manera sencilla para que los usuarios de una empresa puedan tomar me- jores decisiones de negocio. (Malhotra, 2001) describe a BI como un facilitador de conexiones en una nueva forma de organizaci´on, trayendo informaci´on 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 ana´lisis profundo de los datos detallados del negocio, incluyendo base de datos y tecnolog´ıas de aplicacio´n, as´ı como pr´acticas de ana´lisis (Gangadharan y Swami, 2004). BI es t´ecnicamente m´as amplio, lo que potencialmente engloba la gesti´on del conocimiento, la planificacio´n de recursos empresariales, sistemas de apoyo a la toma de decisiones y la miner´ıa 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 reaccio´n necesario para las decisiones empresariales. (Nguyen y cols.,

2005) propone una infraestructura de TI basada en eventos para operar aplicacio- nes de BI que permiten ana´lisis en tiempo real a trav´es de procesos de negocios corporativos, y brindar recomendaciones automa´ticamente 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 reduccio´n del tiempo de acci´on y la interconexio´n 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 cua´ntas mantas fueron vendidas en un mes, las empresas desean entender cua´ntas fueron vendidas en un pa´ıs determinado durante un evento meteorolo´gico. BI proporcio- na una visio´n integrada unificada de las actividades empresariales. Las empresas han construido sistemas de BI que apoyan ana´lisis de negocio y de toma de de- cisiones para ayudarlos a un mejor entendimiento de sus operaciones y competir en el mercado (Gangadharan y Swami, 2004). Algunas innovaciones en tecnolo- g´ıas de almacenamiento de datos esta´n 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 soft- ware con herramientas superiores ofrecen una suite completa de aplicaciones para an´alisis de BI, herramientas y modelos de datos que permiten a una organizacio´n aprovechar su informaci´on. Las herramientas BI facilitan el acceso a un gran volu- men de datos corporativos, y convertir esos datos en informacio´n u´til y procesable que sea consistente a trav´es de la versio´n coherente de la verdad. Las empresas au´n sienten que BI tiene complejidades relacionadas con la tecnolog´ıa y que pue- de usarse solamente por especialistas con conocimientos t´ecnicos, adem´as que los costos de implantaci´on son altos. Las empresas requieren estos an´alisis 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 podr´ıa atender las necesida- des actuales de las empresas. Los datos finalmente son considerados como recursos corporativos en una nueva disciplina. Cualquier sistema transaccional (incluyen- do Sistemas de Planificaci´on de Recursos Empresariales (ERP) y Administracion basada en la relacion con los clientes (CRM)) y cualquier aplicaci´on de apoyo a la decisi´on (incluyendo data warehouses y data marts) son BI, si y solo si fueron desarrollados bajo la proteccio´n y la metodolog´ıa de una iniciativa estrat´egica de toda la Organizacio´n (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 informaci´on para producir la propia inteligen- cia de negocios, y un sistema de informes. Las capacidades de BI incluyen apoyo a la decisi´on, procesamiento anal´ıtico en linea, an´alisis estad´ısticos, predicci´on y miner´ıa de datos. Diferentes sectores como fabricantes, comercios electro´nicos, em- presas de telecomunicaciones, aerol´ıneas, minoristas, sistemas de salud, servicios financieros, bioinform´atica y hoteles utilizan BI para apoyo a clientes, investiga- cio´n de mercado, segmentacio´n, rentabilidad del producto, ana´lisis y distribuci´on

de stock, an´alisis estad´ıstico, informes multidimensionales, detecci´on fraudes, en- tre otros. BI y miner´ıa de datos es un a´rea que esta´ fuertemente influenciado por t´ecnicas estad´ısticas tradicionales, y la mayor´ıa de los m´etodos de miner´ıa de datos revela una fuerte base de m´etodos estad´ısticos y de an´alisis de datos. Algunas de las t´ecnicas tradicionales de miner´ıa de datos incluyen clasificacio´n, agrupacio´n, an´alisis de valores at´ıpicos, patrones secuenciales, ana´lisis de series temporales, la prediccio´n, la regresio´n, ana´lisis de enlaces (asociaciones), y m´etodos multidimen- sionales incluyendo el procesamiento anal´ıtico en l´ınea Procesamiento nal´ıtico en l´ınea (OLAP). Estos pueden clasificarse en una serie de t´ecnicas de miner´ıa de datos, que se clasifican e ilustran en la Tabla 1 (Goebel y Gruenwald, 1999).

|  |  |
| --- | --- |
| TE´ CNICAS | DESCRIPCIO´ N |
| Modelo predictivo | Predecir valor para un atributo espec´ı-  fico del elemento de datos. |
| Caracterizacio´n y miner´ıa de datos des-  criptivo | Distribuci´on, dispersio´n y excepci´on de  datos |
| Asociaci´on, correlaci´on, ana´lisis de la  causalidad (Ana´lisis Link) | Identificar relaci´on entre atributos |
| Clasificacio´n | Determinar a qu´e clase pertenece un  elemento de datos |
| La agrupaci´on y an´alisis de valores at´ı-  picos | Partici´on de un conjunto en clases, con  lo cual elementos con caracter´ısticas si- milares se agrupan |
| Ana´lisis de patrones temporal y se-  cuencial | Tendencia y desviacio´n, patrones se-  cuenciales, frecuencia |
| OLAP(Procesamiento Analitico en Li-  nea) | Herramientas OLAP permiten a los  usuarios analizar distintas dimensiones de datos multidimensionales. Por ejem- plo, proporciona series temporales y puntos de vista de an´alisis de tenden- cias. |
| Modelo de visualizacio´n | Hacer facil la descubierta de cono-  cimiento usando charts, plots, histo- grams y otros medios visuales |
| Ana´lisis Exploratorio de Datos(EDA) | Explorar un conjunto de datos sin una  fuerte dependencia en hipo´tesis o mo- delos; el objetivo es identificar patrones de una manera exploratoria |

Cuadro 2.1: T´ecnicas actuales de BI

En el siguiente cap´ıtulo se presenta una introducci´on a los m´etodos y t´ecnicas de an´alisis exploratorio, y en especial la t´ecnica 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 mu´ltiples or´ıgenes(Marakas, 2003). La mayor parte de la base instalada en todo BI son propietarias de empresas tra- dicionales que han construido sus plataformas alrededor de una capa sem´antica y metadatos, la cual es generalmente accesible solo por herramientas del propio fabricante. La situacio´n actual de cara´cter propietario de la capa sema´ntica tradi- cional de BI fue aceptada y adoptada por m´as organizaciones como un facilitador para ana´lisis .ad hoc.a cambio de una u´nica y confiable versio´n de la realidad, que

puede ser accedida f

cilmente por los usuarios de negocio, ocultando los aspectos

t´ecnicos y la complejidad de las estructuras de datos subyacentes. Sin embargo,

con la adopci´on y r

pido crecimiento de las herramientas de data discovery como

Qlik, Tableau, y Tibco Spotfire, los usuarios buscan cada vez m´as acceso a datos confiables en capas sema´nticas cerradas, y los proveedores de BI se enfrentan a un reto dif´ıcil para seguir siendo relevantes en un mercado que est´a en transici´on. La respuesta de los proveedores tradicionales y la inversio´n significativa hasta la fecha, ha sido la utilizacio´n de sus capas sem´anticas existentes para promover la gesti´on 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 teor´ıa, este enfoque logra un equilibrio entre la facilidad de uso y escalabilidad empresarial. Esto ha mostrado poco ´exito para la mayor´ıa 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 sem´antica que sirve como la base de la mayor´ıa de las plataformas BI tradicionales ha sido ampliamente adoptada por muchas organizaciones a trav´es de los an˜os y ha sido promovida y generalmente aceptada como un componen- te 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, operacio´n y mejora de estas plataformas construidas alrededor de una definida y centralizada capa sema´ntica propietaria.

Este enfoque ha funcionado bien cuando el objetivo fuera una u´nica fuente de datos, en apoyo a sistemas definidos de registros centralizados de informes y gesti´on de dashboards fomentando la consistencia, la gobernanza e integracio´n entre las plataformas de presentacio´n y capas de metadatos. Sin embargo, con el surgimiento y expansio´n de data discovery, el concepto de auto servicio cobr´o preponderancia.

Los usuarios de negocio y an´alisis ahora tienen acceso a un gran rango de herra- mientas que promueven y apoyan el uso auto´nomo sin la participaci´on de TI. Como tal, hay una necesidad emergente para acceder a las reglas de negocio integradas dentro de la capa sema´ntica propietaria de herramientas de BI existentes.

Mientras esto no es posible au´n en la mayor´ıa de los casos hoy en d´ıa, algunos proveedores ya han comenzado a adoptar un enfoque cada vez ma´s abierto para sus capas sema´nticas propietarias, y esto puede llevar a un cambio mayor de mer- cado con el tiempo. La oportunidad probablemente sera´ dictada por el ´exito que los proveedores tradicionales de BI tengan con el desarrollo y la adopcio´n de sus propias ofertas de data discovery, que se ha limitado hasta la fecha.

El acceso abierto a metadatos no es in´edito en BI y en mercados analiticos. Ejemplos incluyen Oracle OBI EE, Microsoft Power, SAP BusinessObjects y, ma´s recientemente, conectividad nativa de Tableau para modelo de datos Birst a trav´es de su capa sema´ntica propietaria.

Oracle BI Enterprise Edition(OBI EE) fue uno de los primeros productos de plataformas BI para tener una capa sem´antica abierta, un vestigio de los or´ıgenes 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 conexi´on ODBC que puede ser consumida por herramientas e interfaces de terceros. Inicial- mente, esto es como Oracle proporciona conectividad a su herramienta interna de informes de produccio´n desarrollada, Oracle BI Publisher. Mas alla´ de eso, pocos clientes son conscientes de esta capacidad, y citan los malos resultados como una raz´on por la que no fue adoptado ampliamente.

La reciente alianza entre Birst y Tableau, establecida en Abril 2015, es el m´as reciente ejemplo de este cambio hacia acceso abierto e integraci´on entre proveedores puros de data discovery y fabricantes tradicionales de BI. Antes del anuncio, Birst no hab´ıa permitido el acceso a su estructura de datos propietaria o capa sema´ntica,

hizo accesible s

lo a de informes, dashboard y capacidades data discovery auto-

contenidas dentro de la cartera Birst. A trav´es de la alianza y desarrollo conjunto, se an˜adi´o una conexi´on con Tableau que permite la conexio´n directa con el medio ambiente Birst. Esto proporciona una mayor flexibilidad y una gama m´as 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 sema´ntica de SAP BusinessObjects (Universe). Este acuerdo permite a los clientes comunes la opci´on 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 solo lectura a herramientas de data discovery

de otros proveedores de BI a la capa sema´ntica propietaria de proveedores de BI tradicionales. Como tal, ellos todav´ıa no apoyan la promocio´n de modelos de datos derivados de herramientas de data discovery en la capa sem´antica como una forma de promover a proveedores independientes que rigen capacidades de data discovery. Esto es, sin embargo, una oportunidad potencial de modernizacio´n que los proveedores tradicionales pueden considerar.

Proveedores de software independientes e integradores de sistemas desarrollar´an nuevas soluciones de capa intermedia que facilitan el acceso a la capa sema´ntica a las herramientas de data discovery a trav´es de servicios web.

Mientras que los clientes prefieren una u´nica solucio´n de sus proveedores titu- lares, ya sea de los proveedores de plataforma tradicional de BI o de proveedor de data discovery, los clientes actualmente encontrara´n ma´s oportunidades para acceder a la capa sema´ntica desde los integradores de software y los proveedores independientes menores.

2.1.3. Framework de An´alisis de Negocio de Gartner

En este trabajo fue adoptado un Framework de An´alisis de Negocio de Gartner, el cual se describe a continuacio´n.

Hay una serie de defectos relacionados en la mayor´ıa de las organizaciones con relaci´on a BI y plataformas analiticas, as´ı como una percepcio´n erro´nea de sus

objetivos y c

mo gestionarlos. Los equipos de BI, especialmente si se encuentran

en el a´rea de TI, creen que:

La plataforma anal´ıtica de negocio debe ser una soluci´on estrechamente inte- grada con pocos componentes, preferentemente de un u´nico proveedor, para entregar una sola versio´n de la verdad para la organizaci´on.

La informaci´on puede ser confiable s

lo si esta´ almacenada en un data wa-

rehouse corporativo y entregada a los consumidores de informacio´n usando artefactos de BI, tales como informes y dashboards.

Las informaciones creadas o manipuladas por los usuarios de negocios inevi- tablemente producir´an discrepancias a trav´es de diferentes ana´lisis, lo que lleva a decisiones equivocadas, generando caos en la organizacio´n a trav´es del tiempo.

La responsabilidad del departamento TI para la gestio´n de informacio´n ter- mina en la capa sema´ntica de BI y en los contenidos orientados a TI. Procesos anal´ıticos orientados al negocio est´an fuera del alcance y no soporta- dos por TI. Hay varios problemas orientados a la informacio´n documentados en el mundo BI y ana´lisis, que obligan a los l´ıderes de BI a seguir estas creencias y desplegar un entorno de BI centralizado y monol´ıtico, que termi- na siendo impuesta a los usuarios, independientemente de su adecuacio´n a las necesidades.

Los proveedores licenciados de BI, a favor de sus propias plataformas, apoyan

este enfoque.

Segu´n Gartner, una plataforma anal´ıtica necesitan evolucionar m´as all´a del pensamiento monol´ıtico. Debe ocurrir una transformaci´on para ofrecer dife- rentes soluciones para las diferentes necesidades del usuario, con un conjunto diverso de niveles de integraci´on, y encontrar un equilibrio entre confianza y agilidad. El prop´osito es ayudar a los usuarios a alcanzar sus objetivos de negocio a trav´es del uso de la tecnolog´ıa apropiada, no para erradicar las solu- ciones de BI orientados al usuario que resuelven parcialmente sus problemas de hoy.

El entorno de BI y el an´alisis resultante tambi´en exigir

cambios en los

procesos de gestio´n de la informacio´n, como atribuir nuevas responsabilidades a diferentes personas en la organizacio´n.

A continuacio´n el framework de ana´lisis de negocios. La figura 2.2 presenta el framework de ana´lisis de negocios de Gartner.

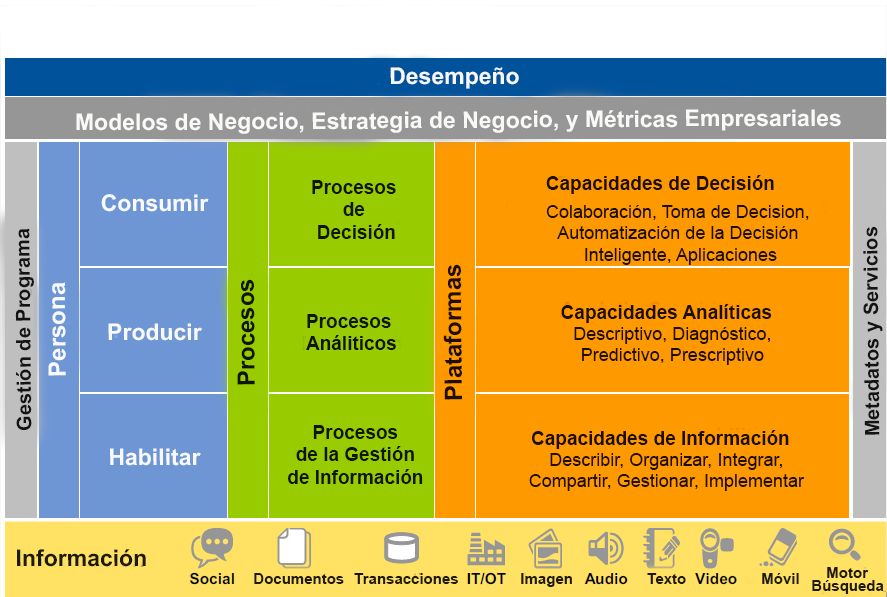


Figura 2.2: Framework Analitico de Negocios Gartner

El framework de an´alisis de Gartner identifica las personas, los procesos y com-

ponentes de la plataforma que apoyan la transformacio´n de la informacio´n en un

mejor rendimiento de la organizaci´on. El uso de esta herramienta est

hecha por

la lectura desde arriba hacia abajo, comenzando con los resultados del negocio y luego descifrando las composiciones anal´ıticas de apoyo y la informaci´on necesaria para alcanzarlos. De acuerdo a las necesidades de los usuarios, la plataforma debe ser redisen˜ada con un amplio conjunto de capacidades t´ecnicas (llenando los va- c´ıos), nuevas responsabilidades y organizaci´on. Centr´andose en las herramientas o normalizacio´n de proveedor por s´ı sola no es la respuesta.

El framework es tambi´en muy u´til 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 organizacio´n muy probablemente tambi´en necesitar´a re-organizar y capacitar a los proveedores y usuarios de BI y ana´lisis. Los usuarios de negocio deben ganar acceso a las herramientas anal´ıticas 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 anal´ıticas con ´enfasis en las plataformas, en particular, el componente de capacidad anal´ıtica, podemos notar cuatro estilos anal´ıticos que se detalladan a continuacio´n (Figura 2.3).

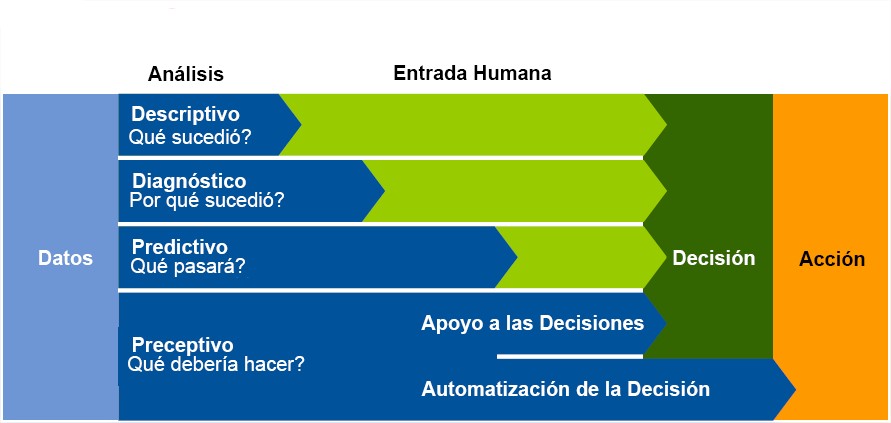


Figura 2.3: Espectro Analitico

Las capacidades anal´ıticas implementadas en organizaciones son a menudo li- mitadas el an´alisis descriptivo, a trav´es de reportes b´asicos y dashboards. Con esto, la pregunta, “Qu´e paso?“ puede ser respondida. Despu´es de conocer “Que,“ ,

lo ma´s probable es que los usuarios tambi´en pregunten, “Por qu´e pas´o?“. Abordar adecuadamente esto requiere mucha ma´s agilidad y ma´s capacidades avanzadas de exploraci´on de la informaci´on. Despliegues de BI tradicionales tienden a tener huecos en esta ´area, pero TI por lo general pasa por alto el impacto de esta pro-

blema´tica y continu´a impulsando el est

ndar del proveedor y sus herramientas no

aptas-para-propo´sito. Como consecuencia, los usuarios recurrir

n a Excel, consul-

tas ad hoc, extracciones de datos y a los equipos de shadow TI para lograr sus metas de an´alisis.

Los l´ıderes de BI deben extender el BI y la plataforma anal´ıtica hasta el ana´lisis de diagno´stico para complementar el ana´lisis descriptivo. Aqu´ı es donde OLAP y los modelos de datos en memoria son utilizadas para proporcionar una navegaci´on facil y r´apida de datos sin una consulta predefinida. Aprovechando mejoras en el nivel de acceso a datos, tambi´en vemos la necesidad de mejorar las capas sem´anticas abstrayendo la complejidad del modelo f´ısico subyacente. Esto puede hacer que sea mucho m´as facil para el descubrimiento de autoservicio sin el cuello de botella de TI que se encuentra en un t´ıpico equipo de BI.

Mas alla´ de la capa de datos, vemos la introduccio´n de nuevas herramientas

de visualizaciones de datos, y aqu´ı es donde el enfoque del r

pido crecimiento de

las herramientas de data discovery se concentran. Pero herramientas tradicionales pueden tambi´en proporcionar mejoras con un mayor enfoque en informes m´as comprensivos(como el ana´lisis de varianza), planificacio´n integrada, dashboards y informes KPI.

A trav´es del tiempo, con un crecimiento a un alto nivel de madurez de ana´- lisis, la organizaci´on deber´ıa moverse dentro del an´alisis predictivo y preceptivo. Esto requiere un incremento significante en los niveles de habilidades del analista de negocios. Modelos predictivos requieren desarrollo y mantenimiento con lo´gicas complejas y reglas de negocio. Ellos incorporan m´etodos sofisticados que pueden tambi´en requerir un entendimiento profundo de la estad´ıstica o investigacio´n ope- racional.

Adema´s, las organizaciones deben darse cuenta de que hay necesidad de mezclar todas estas diferentes t´ecnicas en soluciones integrales en lugar de dejarlos aislados.

Redisen˜ar el BI y la Plataforma Analitica Los l´ıderes de BI deben seguir las herramientas fundamentales descritas anteriormente para as´ı con ´exito redisen˜ar el BI y la plataforma de ana´lisis. Gartner recomienda la instalaci´on de una arqui- tectura por niveles compuesta por:

Portal de Informacio´n. Workbench An´alitico. Laboratorio de datos cient´ıficos.

En la figura 2.4 se presenta la representacio´n de BI en niveles y la plataforma

analitica, la cual puede ser utilizada como una gu´ıa gen´erica que puede ser ajustado de acuerdo a las caracter´ısticas espec´ıficas de la organizaci´on.

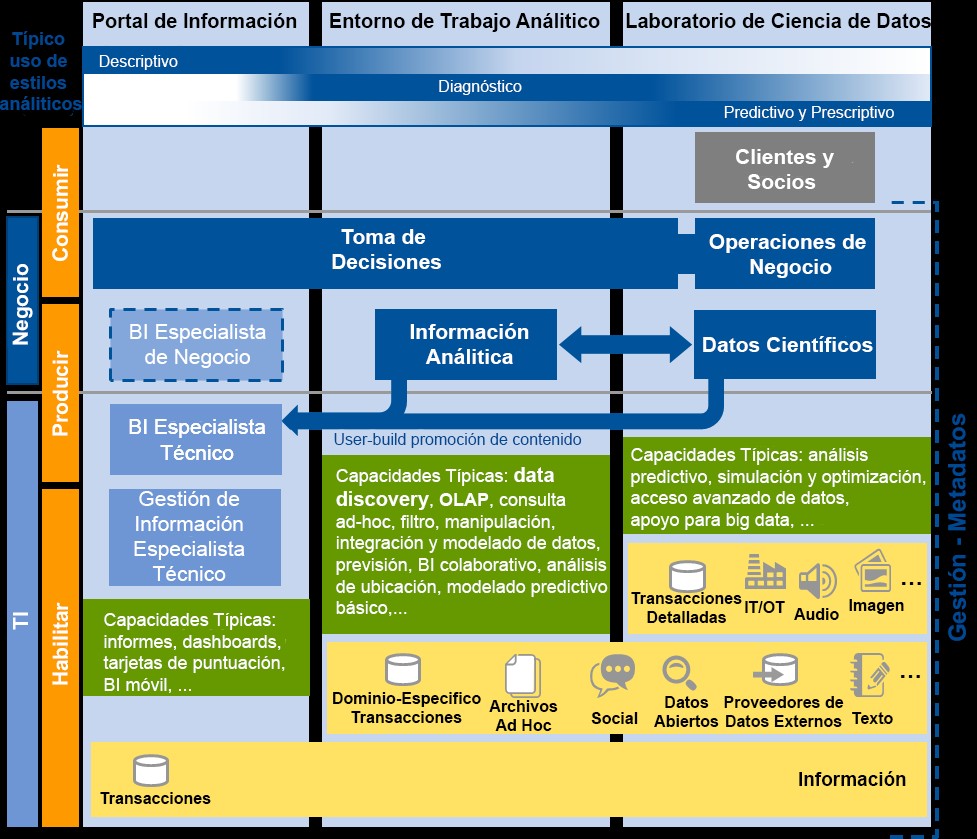


Figura 2.4: Tipico uso de estilos ana´liticos

Para hacer realidad la visio´n de los tres niveles y ser capaz de maximizar sus fortalezas, los l´ıderes de BI necesitan implementar nuevas capacidades t´ecnicas para proporcionar nuevos estilos de ana´lisis, mejorar el uso de las herramientas existentes a trav´es de una mejor integracio´n global, y proporcionar metadatos comunes y gobernanza. Procesos, roles de personas y responsabilidades, son de suma importancia para el ´exito.

Ellos deben ser tratados en conjunto con la plataforma t´ecnica como se describe

en el framework de an´alisis de negocios de Gartner.

Vamos a ampliar cada nivel para entender c conjunto.

2.1.3.1. Portal de informacion

mo integrar y aprovecharlos en

Seguir de cerca las caracter´ısticas de sistemas de registros desde el Pace Layer

Model. El portal de informacio´n es el a´rea de trabajo donde los usuarios de negocio

pueden encontrar r

pida y f´acilmente las m´etricas clave de confianza con la cual la

organizaci´on mide su rendimiento. Por lo general hecho de capacidades de informes y dashboard que proporciona contenido para los consumidores de informaci´on. Sus productos son el resultado de un proceso de desarrollo formal que abarca que un usuario de negocios establezca requisitos y un especialista t´ecnico(t´ıpicamente de TI, pero cada vez m´as de negocio) los implemente.

Capacidades t´ıpicas de la plataforma: Informes.

Dashboards.

Integracio´n Microsoft Office. BI m´ovil.

Ana´lisis integrada.

Ejemplo de herramientas y proveedores: SAP BusinessObjects.

IBM Cognos. Oracle BI.

Microsoft Reporting Services. MicroStrategy.

Information Builders WebFocus.

2.1.3.2. Workbench anal´ıtico

El workbench anal´ıtico es el ´area de trabajo usado para investigar tendencias en indicadores de confianza o detectar patrones en otros conjuntos de datos — desde mu´ltiples fuentes –. que pueden convertirse en oportunidades o riesgos. Es una capa ´agil para explorar informacio´n y tener acceso a un amplia gama de fuentes de datos, con limitado o ningu´n apoyo de expertos t´ecnicos. Los conjuntos de herramientas deber´ıan incluir una herramienta de data discovery (descubrimiento de datos) y un nu´mero de otras capacidades para ayudar a los usuarios de negocios a extraer valor de la informacio´n de forma auto´noma.

En el espectro de ana´lisis, el workbench es capaz de proporcionar ana´lisis des-

criptivo ,pero por lo general se centrar

en el ana´lisis de diagno´stico. En algunos

casos – es decir, a trav´es del uso de herramientas de data discovery ma´s centrado al ana´lisis. – Puede extender a un nivel b´asico de ana´lisis predictivo y ganar´a el modelado de datos y capacidades anal´ıticas m´as avanzadas en el futuro.

Capacidades t´ıpicas de la plataforma: Data discovery.

Informes/consultas Ad hoc. Inteligencia geoespacial y localizacio´n. Ana´lisis integrado avanzado.

OLAP.

Mashup de datos y modelado de usuarios de negocios. Colaboraci´on.

Filtrado y manipulacio´n 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.

2.1.3.3. Laboratorio de datos cient´ıficos

El laboratorio cient´ıfico 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 organizaci´on.

Un amplio conjunto de capacidades t´ecnicas es esperado y a menudo propor- cionado por herramientas especializadas con integracio´n TI m´ınima, 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 inversio´n en el portal de informacio´n. Los usuarios son cualificados y experimentados, a menudo ma´s que los expertos t´ecnicos en TI. Sus conjuntos de herramientas incluyen capacidades de data mining, predicciones y otras herramientas de estad´ısticas y an´alisis complejos.

Capacidades t´ıpicas de la plataforma: Acceso avanzado de datos.

Soporte para fuentes de big data. Ana´lisis descriptivo avanzado. Ana´lisis predictivo.

Predicci´on. Optimizacio´n. Simulaci´on.

Otros an´alisis avanzado.

Aunque no capacidades BI, Hadoop y bases de datos NoSQL tambi´en deben ser referenciados aqu´ı.

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 caracter´ısticas de BI en niveles y la plataforma anal´ıtica. La si- guiente tabla resume las caracter´ısticas que posee cada una de las tres capas. Los l´ıderes de BI deber´ıan debera´n intentar entender los huecos en su plataforma ana- l´ıtica, y cambiar sus estrategias en consecuencia.



Figura 2.5: Caracter´ısticas de BI en niveles y Plataforma Anal´ıtica

23

CAPITULO 3

3.1. Selecci´on de la herramienta para Data Dis- covery

Fueron analizados los estudios de Gartner(Herschel, Linden, y Kart, 2015)(Sallam,

2015) para la seleccio´n de la mejor herramienta que se adecue a los criterios nece-

sarios para ser utilizado en esta tesis. Estos documentos realizan un an´alisis de las

mejores herramientas del mercado, en un a´rea de conocimiento. A continuaci´on se

presenta el Cuadrante M´agico de Gartner, para herramientas de BI y Analytics.

3.1.1. Cuadrante M´agico de Gartner



Figura 3.1: Cuadrante M´agico para BI y Plataformas Anal´ıticas

Los l´ıderes del mercado se encuentran siempre en el cuadrante superior derecho. Se puede observar una amplia diferencia entre Tableau y los dema´s l´ıderes.

En las siguientes figuras se presenta el an´alisis de Gartner que evalu´a las capa- cidades cr´ıticas que debe tener una herramienta de BI y Analytics, para adecuarse a las necesidades del mercado.

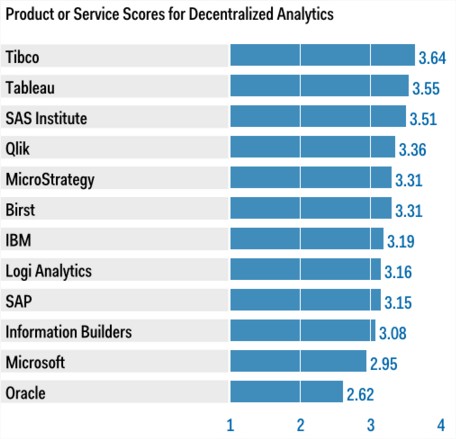


Figura 3.2: Puntuaciones de producto o servicio para ana´lisis descentralizado

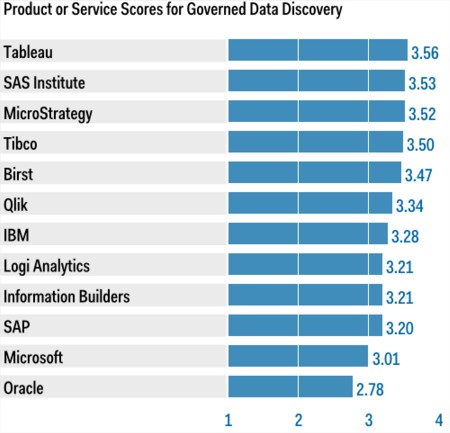


Figura 3.3: Puntuaciones de producto o servicio guiados por Data Discovery

Los valores posibles van del 1 al 5, conforme la siguiente evaluacio´n:

1. Pobre o ausente: la mayor´ıa 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 posici´on fuerte en capacidad de ejecuci´on (producto/servicio, su oferta, ejecucio´n de ventas, marketing, experiencia del cliente) en el eje de l´ıderes del cuadrante. Esta herramienta fue la que mejor se adecu´o a las necesidades del trabajo de Tesis, dado que cuenta con una versi´on pu´blica para la construcci´on y publicacio´n de dashboards, adem´as de la facilidad de uso que nos proporciona. Tableau Desktop, la cual se basa en tecnolog´ıa drag and drop (arrastrar y soltar) permite analizar datos r´apidamente y permite ver los cambios en tiempo real sin necesidad de codificacio´n, de esta manera, posibilita a un usuario con escasos conocimientos t´ecnicos, poder utilizarlo de igual manera.

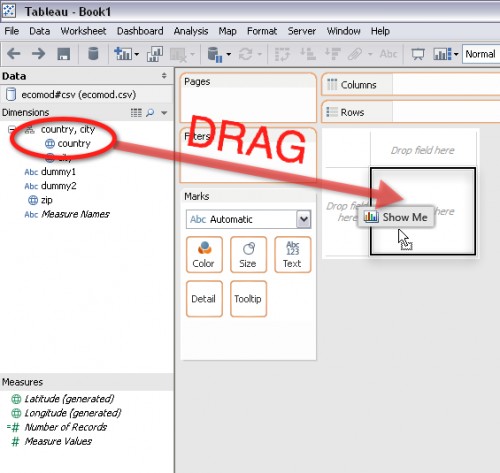


Figura 3.4: Arrastre el campo pa´ıs para el campo desplegable sen˜alado.

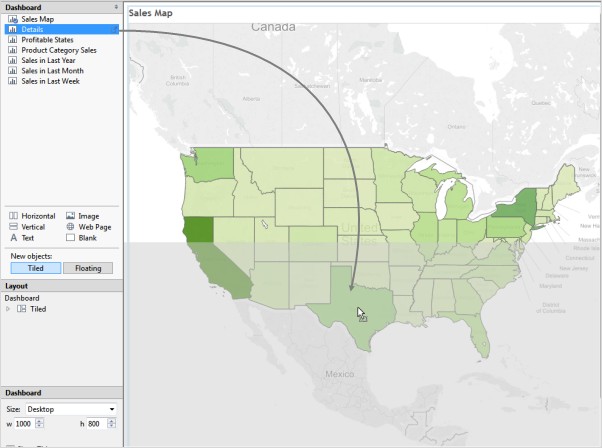


Figura 3.5: Arrastrar hojas de trabajo al dashboard.

De una forma a´gil el usuario puede conectarse a diversas fuentes de datos y crear paneles interactivos, conectando entre s´ı los diferentes componentes (tipo de

gr fico) que proporciona la herramienta. La herramienta permite utilizar compo-

nentes 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 organiza-

ciones para comprender ra´pidamente diferentes aspectos del negocio. Tambi´en se

puede utilizar para realizar proyecciones o tendencias, la cual Tableau nos ofrece

de manera automa´tica.

3.2. Aplicaci´on de Data Discovery a datos de ins- tituciones del Estado

Generalmente las organizaciones no logran comprender en su totalidad los da- tos que generan. La consecuencia de no comprender esos datos puede resultar en la mala toma de decisio´n, lo cual podr´ıa ocasionar un gran impacto negativo a la organizacio´n. La informaci´on es considerada como uno de los recursos ma´s im- portantes en una organizacio´n, y en base a esta informacio´n, se puede obtener conocimiento que podr´ıa 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 informacio´n de inter´es para la organizacio´n.

3.2.1. Datos de la ANDE y de la DGEEC

Se cuenta con datos de consumo de energ´ıa el´ectrica, facturaciones, grupos de consumo (residencial, industrial, exportacio´n, comercial, gubernamental y otros), por an˜o (2000-2014), por departamento y distrito. Estos datos fueron solicitados

formalmente a la instituci´on por medio de la Facultad de Ciencias y Tecnolog´ıa de la Universidad Cato´lica, a la cual tuvimos una respuesta favorable.

3.2.2. Dashboard de control / monitoramiento

En esta secci´on 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 t´ecnicas 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 an˜o (Ej: Si al cerrar el an˜o 2014, la cantidad de clientes fue de 1.000.000 y en el an˜o 2015 aumento´

100.000, esto quiere decir que en el an˜o 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 an˜o 2016 la ANDE cierra con un total de

1.000.000 de clientes, su crecimiento ser a 10 % menor al an˜o anterior. La formula empleada (ver Fo´rmula abajo), donde n es el an˜o actual y n-1 el an˜o anterior, PIB indica la cantidad de clientes que posee la ANDE .

P I Bn − P I Bn−1

tn =

P I B

n−1

× 100

3.2.2.1. Dashboard - Clientes Facturados vs Crecimiento Poblacional

En el dashboard de la Figura 3.6 son utilizados datos histo´ricos de la po- blaci´on, prove´ıdos por la DGEEC, y datos de clientes, tales como el consumo y facturacio´n, prove´ıdos por la ANDE. Es importante notar en la figura, que inde- pendiente a que la tasa de crecimiento poblacional se mantenga casi constante, la tasa de crecimiento del consumo se eleva de forma pronunciada. Adem´as, puede ser muy relevante la informaci´on de proyeccio´n del consumo para los pro´ximos an˜os, para elaborar una planificaci´on en la ampliacio´n de la capacidad de transmisi´on o distribucio´n de energ´ıa.

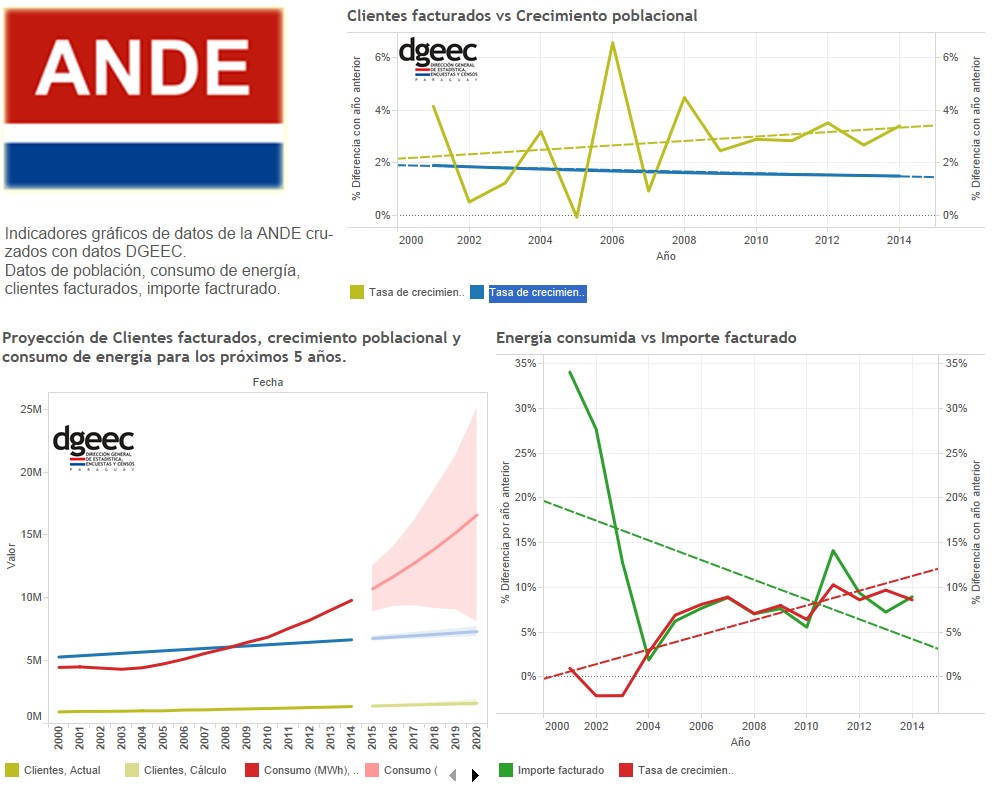


Figura 3.6: Clientes Facturados vs Crecimiento Poblacional

En la Figura 3.7 se presentan dos tasas de crecimiento: de Clientes de la ANDE y de la Poblacio´n del Paraguay. Es importante notar que el ritmo de crecimiento de la poblaci´on disminuye con el tiempo, pero no as´ı el ritmo de consumo de electri- cidad. No existe una relacio´n directa entre estas dos tasas por el momento, lo cual puede deberse a factores tales como instalaci´on de mayor cantidad de dispositivos el´ectricos en residencias, o bien la instalacio´n de ma´s 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 podr´ıa ayudar a planificar una mayor inversio´n en la red de electricidad para aquellas zonas donde se registran tendencias mayores al consumo.

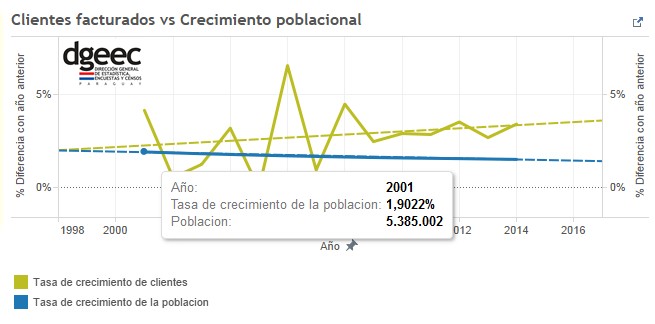


Figura 3.7: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.8 observamos que la cantidad de la poblaci´on en el an˜o 2002 cerro´ con un total de 5.484.610, la cual su crecimiento fue del 1,8497 % que equivale a 99608.

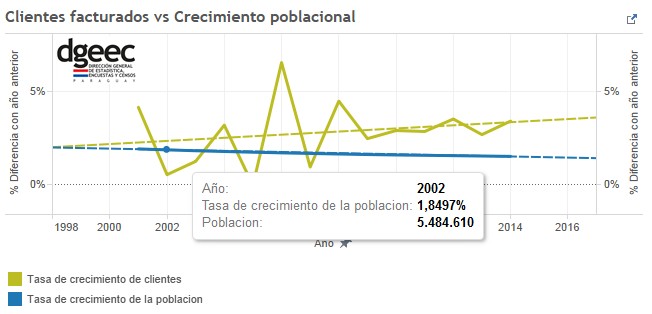


Figura 3.8: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.9 se muestran valores que representan el porcentaje del crecimien- to de los clientes de la ANDE. Como podemos ver, hay an˜os en que el aumento es muy evidente (2004,2006,2008) y hay an˜os en que este es m´ınimo(2002,2005,2007).

Las l´ıneas discontinuas representan las tendencias de ambos puntos. Por ejemplo, la cantidad de clientes en el an˜o 2001 fue de 959.580, la cual aument´o el 4.1 % respecto al an˜o anterior.

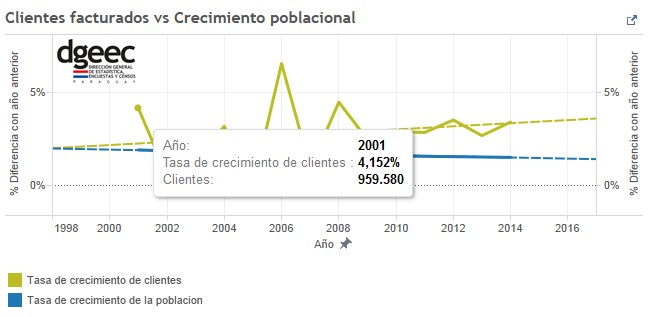


Figura 3.9: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.10, se muestra la cantidad de clientes correspondiente al an˜o

2002, vemos que ascendi´o a 964.449 con un aumento de 4.869, que corresponde a

un incremento del 0.5 % respecto al an˜o 2001. Sin embargo en la figura 3.11, en

el an˜o 2003 el incremento fue de 1.2 %, la cual representa a un aumento de ma´s

que el doble del an˜o anterior llegando a aumentar 11830 clientes.



Figura 3.10: Clientes Facturados vs Crecimiento Poblacional

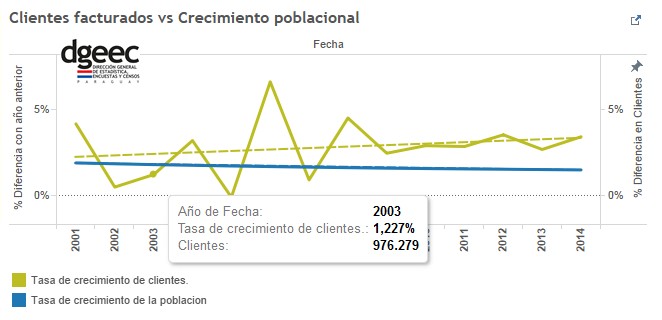


Figura 3.11: Clientes Facturados vs Crecimiento Poblacional

En la figura 3.12 se puede observar con ma´s detalle la proyeccio´n (forecasting) del consumo, crecimiento poblacional y cantidad de clientes. Se puede notar que existe una mayor probabilidad que en promedio, el consumo de electricidad au- mente de forma considerable hasta el 2020 (l´ınea roja). Adem´as, este aumento es mayor en proporcio´n al aumento de la poblacio´n y a la cantidad de clientes.

Esto demuestra claramente que el consumo aumenta cada vez ma´s, indepen- diente que se d´e un aumento en la misma proporcio´n de clientes o poblaci´on. Esto

puede ocurrir a causa de varios factores, entre ellos, el aumento de dispositivos el´ectricos en las residencias, debido al aumento en la capacidad adquisitiva de las personas. Tambi´en se puede dar a causa de un aumento en la cantidad de indus- trias.

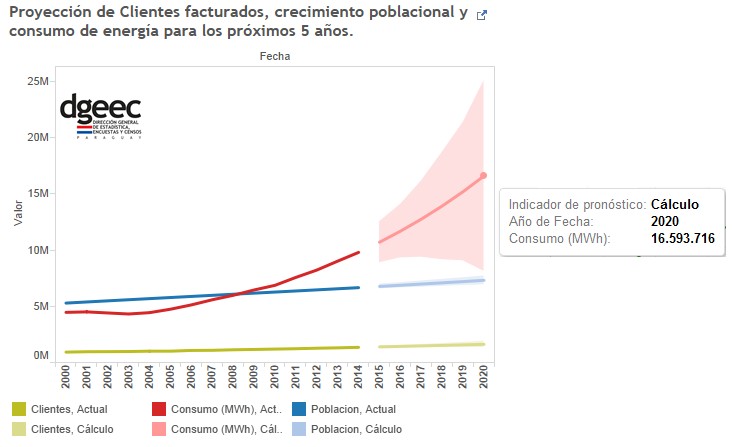


Figura 3.12: Proyecci´on de Clientes Facturados Crecimiento Poblacional Y Consu- mo De Energ´ıa para los Pr´oximos 5 An˜os

En el tercer y u´ltimo gra´fico de este panel, se muestra el porcentaje del creci- miento anual de los importes facturados y consumo de energ´ıa. Se puede observar que la facturacio´n de la ANDE, en general es proporcional al consumo de energ´ıa el´ectrica, excepto en el an˜o 2011 (Figura 3.13), en la cual el importe facturado fue superior al aumento del consumo de energ´ıa. Sin embargo, en el an˜o 2013 (Figura

3.14) se registra nuevamente un aumento en relacio´n a an˜os anteriores.

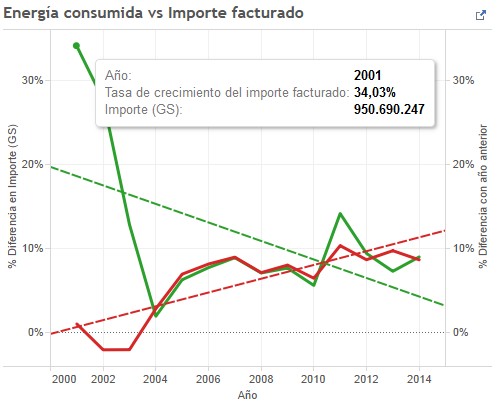


Figura 3.13: Energ´ıa Consumida vs Importe Facturado

36

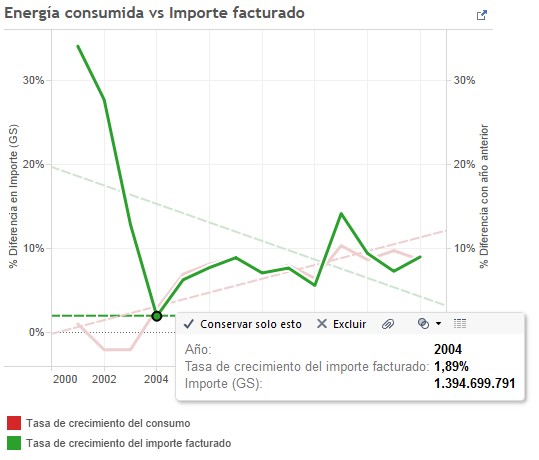


Figura 3.14: Energ´ıa Consumida vs Importe Facturado

3.2.2.2. Panel estad´ıstico de consumo de electricidad por sector

En la figura 3.15 se presenta un gr´afico con el consumo de energ´ıa hist

ri-

ca, que abarca desde el an˜o 1990 hasta 2014. Estos datos esta´n clasificados por

los siguientes criterios de la ANDE: Alumbrado Pu´blico, Comercial, Exportacio´n,

Gubernamental, Industrial y Residencial. En este gr´afico se puede apreciar que

el mayor consumo de energ´ıa se encuentra en el sector Residencial. Sin embargo,

los valores de Exportacio´n de energ´ıa 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 energ´ıa en el

pa´ıs, por lo tanto disminuye la energ´ıa disponible para exportacio´n.



Figura 3.15: Estad´ıstica de consumo de electricidad por sector(1990-2014)

La Figura 3.16 presenta las informaciones de facturacio´n tambi´en clasificados por sector, con el recurso de filtros por an˜o. Es importante destacar un factor resaltante: aunque energ´ıa exportada disminuyo´, el valor facturado por energ´ıa vendida al exterior aument´o. Es probable que esto se haya debido al aumento de la tarifa de energ´ıa vendida, lo cual beneficia al pa´ıs.



Figura 3.16: Importe facturado por an˜o y sector(1990-2014)

3.2.2.3. Panel comparativo de Tasa de crecimiento y el Consumo de energ´ıa

En la figura 3.17 se presenta un panel comparativo entre la tasa de crecimiento de clientes y consumo de energ´ıa. Tambi´en se cuenta con un mapa para filtrar por regi´on del pa´ıs.

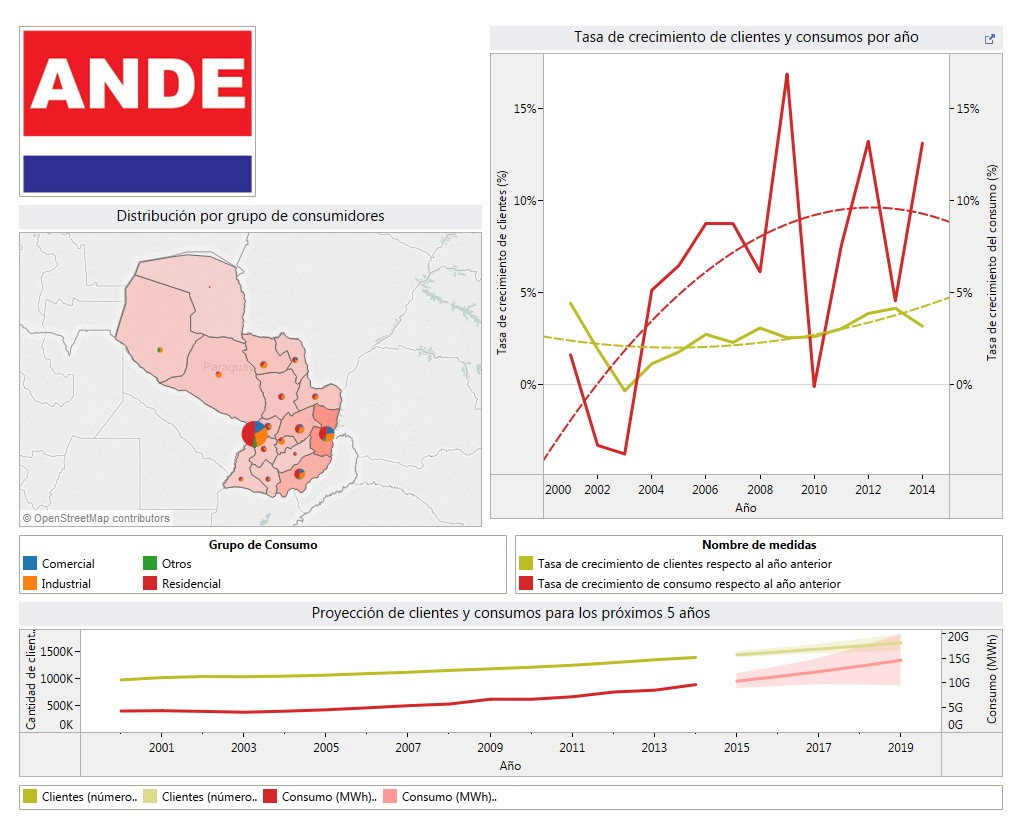


Figura 3.17: Tasa de crecimiento vs Consumo de energ´ıa

En el gr

fico 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.

As´ı se puede apreciar, que existe una tendencia de crecimiento sostenido durante

el tiempo de clientes. Sin embargo, la tendencia que el consumo crezca es mayor

al de clientes.

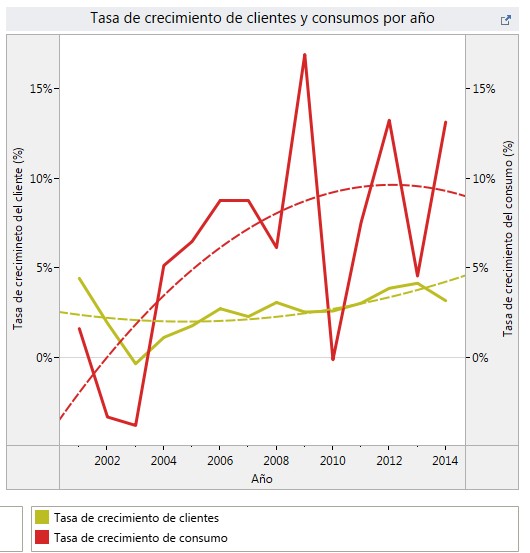


Figura 3.18: Tasa de crecimiento vs Consumo de energ´ıa, filtrado por el departa- mento Alto Parana´

Debajo se presenta un cuadro con la l´ınea de cada valor (Figura 3.19), de creci- miento y consumo. Con el uso de un recurso disponible que cuenta la herramienta escogida Tableau, es posible realizar ana´lisis predictivo (forecasting) del crecimien- to y consumo. Tableau utiliza un algoritmo llamado Suavizado Exponencial, muy

conocido en el ´area de Matem´aticas Estad´ısticas. En este gr

fico se puede notar

que existe una mayor probabilidad que en los pro´ximos an˜os aumente considera-

blemente 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.

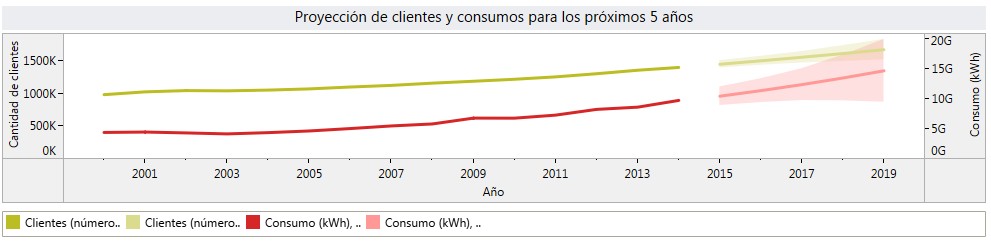


Figura 3.19: Proyeccio´n de clientes y consumos para los pro´ximos 5 an˜os

3.2.2.4. Panel de Tasa de crecimiento poblacional y Consumo de ener- g´ıa anual

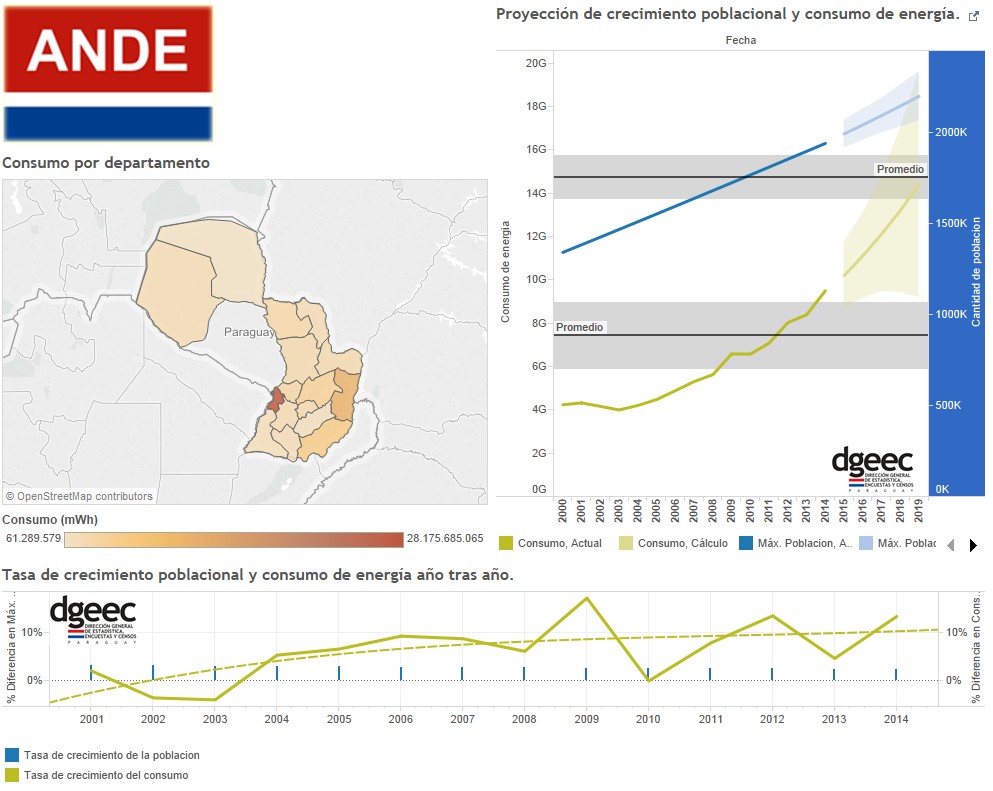


Figura 3.20: Panel de Tasa de crecimiento poblacional y Consumo de energ´ıa anual

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 ma´s oscuro el consumo de energ´ıa es mayor, y si el color es ma´s claro, el consumo es menor. Se puede notar que los departamentos Central y Alto Paran´a son los que tiene mayor consumo de energ´ıa. Este tipo de gr´afico es muy u´til cuando se desea analizar informacio´n 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 dema´s gr´aficos tambi´en se actualizan filtrando el departamento seleccionando.

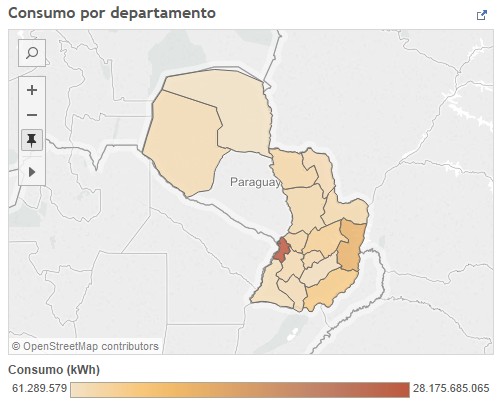


Figura 3.21: Consumo por departamento

Los gr´aficos situados a la derecha y debajo del mapa presentan la tasa de crecimiento poblacional y de consumo de energ´ıa. Es importante notar que la tasa de crecimiento de la poblacio´n es casi constante, esto es, no tiene una gran variacio´n en el tiempo, ni una tendencia a alejarse de la media. Sin embargo, la tasa de crecimiento de consumo tiene una l´ınea de tendencia a crecer durante el

tiempo. As´ı puede apreciarse que, el consumo de energ´ıa no tiene una relaci´on de proporcionalidad con respecto al crecimiento de la poblaci´on.

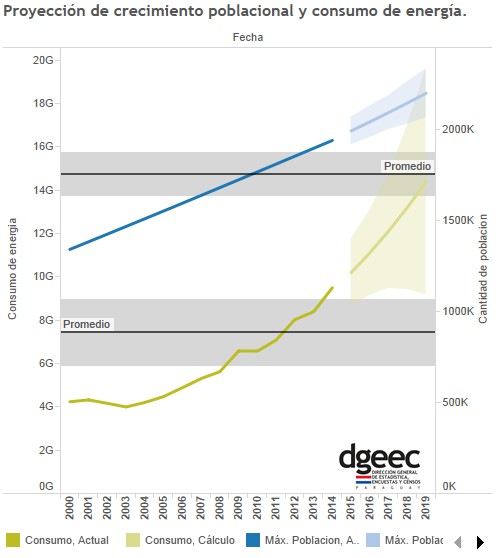


Figura 3.22: Proyeccio´n de crecimiento poblacional y consumo ee energ´ıa

En este gr´afico, se muestra la misma informacio´n que el gra´fico anterior pero con diferente perspectiva, en este caso se calcula el porcentaje de crecimiento anual tanto de la poblacio´n, as´ı como del consumo.



Figura 3.23: Tasa de crecimiento poblacional y consumo de energ´ıa an˜os tras an˜os

45

CAPITULO 4

4.1. Marco Metodol´ogico

4.1.1. Alcance

Aplicaremos las t´ecnicas de Data Discovery a los datos de dos instituciones del estado, espec´ıficamente la ANDE y DGEEC, donde demostraremos que con datos de calidad podr´ıamos detectar oportunidades que nos faciliten la toma de decisiones en la institucio´n. Utilizaremos conjuntos de datos de las instituciones mencionadas ma´s arriba para este fin.

4.1.2. Enfoque

El enfoque que utilizamos es el cuantitativo, que por lo comu´n, utiliza la reco- lecci´on y el ana´lisis de datos para contestar preguntas de investigaci´on y probar hipo´tesis establecidas previamente, y conf´ıa en la medicio´n num´erica, el conteo, y en el uso de la estad´ıstica para intentar establecer con exactitud patrones en una poblacio´n. (por ejemplo un censo es un enfoque cuantitativo del estudio demogr´a- fico de la poblacio´n de un pa´ıs).

4.1.3. T´ecnica e Instrumentos de recoleccion de datos

La t´ecnica aplicada en este trabajo en la recoleccio´n de datos fue la investigaci´on de documentos cient´ıficos procedentes de publicaciones de empresas pioneras en Data Discovery y de expertos en el a´rea.

CAPITULO 5

5.1. Conclusiones y Traba jos futuros

Este trabajo fue dividido en las siguientes partes:

- Cap´ıtulo 1: en este cap´ıtulo fue presentada una introduccio´n a los objetivos

de este trabajo.

- Cap´ıtulo 2: Marco Teorico: esta seccio´n desarrolla el estado del arte en el

a´rea de BI y principalmente el de Data Discovery, t´ecnica aplicada en este trabajo.

- Cap´ıtulo 3: En esta seccio´n se presenta la evaluacio´n t´ecnica realizada para

la seleccio´n de la herramienta Tableau. Adem´as la aplicaci´on de t´ecnicas de Data

Discovery a los datos de la ANDE y la DGEEC. Tambi´en se presentan los productos

construidos en este trabajo, para el an´alisis de las informaciones, y las proyecciones

realizadas para los pro´ximos an˜os.

- Cap´ıtulo 4: En esta secci´on se presenta el marco metodol´ogico utilizado

para el desarrollo de esta tesis, incluyendo el alcance, el enfoque y la t´ecnica e

instrumentos de recoleccio´n de datos.

Para el desarrollo de este trabajo se obtubieron datos de la ANDE y la DGEEC.

Con estos datos fue posible aplicar t´ecnicas de Data Discovery realizar un analisis

de las informaciones, cruzarlos, georeferenciarlos, encontrar l´ıneas de tendencias

y prono´sticos de crecimiento a futuro tanto del consumo de energ´ıa, de clientes

de la ANDE y de la poblaci´on del pa´ıs. En cada caso se presenta un an´alisis que

demuestra con gr

ficos intuitivos que en ciertas ocasiones no existe una relacion

proporcional entra algunas dimensiones. Sin embargo, teniendo en cuenta los re-

sultados 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 energ´ıa del pa´ıs, disminuy´o la cantidad de energ´ıa exportada.

3. Aunque se tuvo una disminucio´n en la energ´ıa exportada, se obtuvo un cre- cimiento en el valor facturado. Esto demuestra una mejor´ıa en el precio de venta de la energ´ıa al Brasil (Itaipu) o Argentina (Yacyreta).

4. Segu´n el prono´stico de crecimiento para los pro´ximos an˜os, el consumo de energ´ıa tendra´ un crecimiento mayor al de la cantidad de clientes, y la po- blaci´on del pa´ıs.

5. La tasa de crecimiento de la poblacio´n se mantiene constante durante el tiempo, esto es, la poblacio´n crece a una tasa sostenida. Sin embargo, la tasa de aumento en el consumo de energ´ıa es considerablemente mayor, y

47

demuestra un aumento abrupto para los pr´oximos an˜os, independiente a la tasa de aumento de la poblacio´n y de nuevos clientes de la ANDE.

Teniendo en cuenta estas cuestiones, este trabajo puede servir de herramienta para la Planificaci´on en la Inversi´on en la capacidad de transmisio´n y distribucio´n de electricidad dentro del territorio del pa´ıs, para los pr´oximos an˜os, adema´s de ayudar con indicadores para montar una estrategia de exportaci´on para replantear las condiciones actuales de venta de la energ´ıa 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 , 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.

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 inte- lligence and analytics platforms. Gartner Research. Gartner Inc, 1 .

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

51