

UNIVERSIDAD PERUANA UNIÓN

FACULTAD DE INGENIERÍA Y ARQUITECTURA

Escuela Profesional Ingeniería de Sistemas



Una Institución Adventista

**Implementación de una plataforma de Business Intelligence para
la toma de decisiones en el Centro de Recursos para el
Aprendizaje y la Investigación (CRAI) de la
Universidad Peruana Unión**

Por:

Narda Naomi Lanuza Bustamante

Asesor:

Dr. Jorge Alejandro Sánchez Garcés

Juliaca, mayo de 2019

DECLARACION JURADA DE AUTORIA DEL INFORME DE TESIS

Dr. Jorge Alejandro Sánchez Garcés, de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas, de la Universidad Peruana Unión.

DECLARO:

Que el presente informe de investigación titulado: "IMPLEMENTACIÓN DE UNA PLATAFORMA DE BUSINESS INTELLIGENCE PARA LA TOMA DE DECISIONES EN EL CENTRO DE RECURSOS PARA EL APRENDIZAJE Y LA INVESTIGACIÓN (CRAI) DE LA UNIVERSIDAD PERUANA UNIÓN" constituye la memoria que presenta la bachiller Narda Naomi Lanuza Bustamante para aspirar al título Profesional de Ingeniero de Sistemas ha sido realizada en la Universidad Peruana Unión bajo mi dirección.

Las opiniones y declaraciones en este informe son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente declaración en Juliaca a los trece días del mes de mayo del año dos mil diecinueve

Dr. Jorge Alejandro Sánchez Garcés

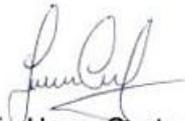


Implementación de una plataforma de Business Intelligence para la
toma de decisiones en el Centro de Recursos para el Aprendizaje y la
Investigación (CRAI) de la Universidad Peruana Unión

TESIS

Presentada para optar el título profesional de Ingeniero de Sistemas

JURADO CALIFICADOR



Mg. Lennin Henry Centurión Julca
Presidente



Mg. Roel Dante Gómez Apaza
Secretario



Mg. Nilton Omar Santillán Aching
Vocal



Mg. Esteban Tocto Cano
Vocal



Dr. Jorge Alejandro Sánchez Garcés
Asesor

Juliaca, 13 de mayo de 2019

DEDICATORIA

A Dios por sus cuidados y bendiciones en cada etapa de mi vida. A mi esposo Daniel por su amor, apoyo constante e incondicional durante todo este tiempo, motivándome a concluir este reto. A mis padres: Francisco Lanuza, Deodata Bustamante, Lorenzo López y Elizabeth Venancio, quienes siempre me apoyaron en todo momento. A mis hermanas: Almendra, Zarela, Yonild, Lizbeth, Rina y Raquel, quienes tuvieron presente en sus oraciones el desafío de concluir el proceso de tesis. A mi mejor amiga: Ana Rivera, futura médico neurocirujana, para motivarla a concluir su investigación en el campo de la medicina. A mis futuros hijos, que cuando sean grandes y lean esta tesis puedan ser motivados a ser investigadores.

AGRADECIMIENTOS

Al Dr. Jorge Alejandro Sánchez Garcés, por asesorarme durante toda la investigación a tiempo y fuera de tiempo, mostrándome que se puede experimentar siguiendo la intuición, de manera que mi mentalidad se abre y no se encierra en un concepto cuadrado. Dios lo bendiga a usted y su familia, por sus oraciones elevadas al cielo por sus asesorados, su esfuerzo y dedicación puestos en la presente investigación.

Al Centro de Recursos para el Aprendizaje y la Investigación (CRAI) por permitirme hacer esta investigación.

A la Universidad Peruana Unión, mi alma máter, a mis profesores que siempre tuvieron disposición de ayudarme y compartir sus conocimientos conmigo.

TABLA DE CONTENIDO

RESUMEN	xii
ABSTRACT	xiii
CAPÍTULO I. EL PROBLEMA	14
1.1. Descripción de la situación problemática.....	14
1.2. Justificación	17
1.3. Presunción filosófica	19
1.4. Objetivos de la investigación.....	20
1.4.1. Objetivo general:	20
1.4.2. Objetivos específicos:.....	20
CAPÍTULO II. BASES TEÓRICAS	21
2.1. Revisión de la literatura.....	21
2.1.1. Metodología de Ralph Kimball vs Bill Inmon	21
2.1.2. Herramientas a utilizar:	22
2.2. Marco teórico.....	41
2.2.1. Investigaciones relacionadas	41
2.2.2. Business Intelligence	43
2.2.3. Toma de decisiones	44
2.2.4. Metodología de Ralph Kimball	48
2.2.5. Matriz de similaridad.....	48
CAPÍTULO III. METODOLOGÍA	50
3.1. Descripción del lugar de ejecución.....	50
3.2. Materiales e insumos	50
3.3. Metodología.....	50
3.3.1. Tipo de investigación	50
3.3.2. Arquitectura de solución.....	51
3.4. Metodología de Ralph Kimball	55

CAPÍTULO IV. RESULTADOS Y DISCUSIÓN	81
4.1. Resultados.....	81
4.1.1. Resultado 1	81
4.1.2. Resultado 2	81
4.1.3. Resultado 3	82
4.1.4. Resultado 4	83
4.1.5. Resultado 5	84
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES	85
5.1. Conclusión 1	85
5.2. Conclusión 2	85
5.3. Conclusión 3	85
REFERENCIAS BIBLIOGRÁFICAS	87
ANEXOS	93

ÍNDICE DE TABLAS

Tabla 1. Demora en obtener reportes del sistema.....	16
Tabla 2. Características de las metodologías.....	21
Tabla 3. Herramientas para la transformación y limpieza de datos.....	25
Tabla 4. Herramientas para la visualización de los reportes	34
Tabla 5. Características y ventajas de Spyder	35
Tabla 6. Divisiones principales del sistema de clasificación decimal Dewey.....	39
Tabla 7. Materiales e insumos	50
Tabla 8. Cronograma del proyecto	56
Tabla 9. Requerimientos funcionales	58
Tabla 10. Requerimientos No Funcionales.....	59
Tabla 11. Softwares de instalación	59
Tabla 12. Reportes de visualización de datos.....	64
Tabla 13. Top de los 10 libros más recomendados	84

ÍNDICE DE FIGURAS

Figura 1. Cuadrante mágico para plataformas de análisis e inteligencia de negocios.....	26
Figura 2. Esquema del proceso ETL	36
Figura 3. Vista de la página de inicio de la tabla de Málaga.....	40
Figura 4. Proceso de toma de decisiones	46
Figura 5. Arquitectura de solución BI de Escobar & Gaete	51
Figura 6. Arquitectura de solución de BI de Contel	52
Figura 7. Arquitectura de solución BI de Peña.....	52
Figura 8. Arquitectura de solución de la plataforma BI	53
Figura 9. Arquitectura de solución del sistema de recomendación	54
Figura 10. Metodología de Ralph Kimball.....	55
Figura 11. Estructura de Desglose de Trabajo del proyecto.....	57
Figura 12. Proceso de préstamo de libro	58
Figura 13. Vista del panel de exportación de informes SirsiDynix Symphony	60
Figura 14. Data histórica en bruto	60
Figura 15. Función Excel CONTAR.SI.CONJUNTO	61
Figura 16. Función Excel TRUNCAR	61
Figura 17. Función Excel SI.....	61
Figura 18. Función Excel SUMAR.SI.CONJUNTO.....	62
Figura 19. Menú principal del sistema académico	62
Figura 20. Vista de la búsqueda de los datos del alumno.....	63
Figura 21. Vista de la búsqueda de datos del docente	63
Figura 22. Base de datos limpia	63
Figura 23. Conversión de archivo a xlsx a csv en Pentaho Data Integration (Kettle).....	64
Figura 24. Panel de inicio de Tableau	65
Figura 25. Carga del archivo limpio al Tableau	66
Figura 26. Panel de dimensiones y medidas del proyecto.....	67
Figura 27. Vista de configuración de dimensiones y medidas	68
Figura 28. Elaboración del reporte de préstamos de las carreras según el mes del año	68
Figura 29. Configuración del reporte de cantidad de préstamos según total de cada mes del año	69
Figura 30. Elaboración del reporte de préstamos de las carreras según el mes del año	70
Figura 31. Panel de análisis de Tableau	71

Figura 32. Configuración del reporte de pronóstico de cantidad de préstamos según el mes del año	71
Figura 33. Elaboración del reporte de pronóstico de préstamos de las carreras según el mes del año	72
Figura 34. Configuración del reporte del promedio de cantidad de préstamos hechos según la hora del día y año.....	73
Figura 35. Elaboración del reporte del promedio de cantidad de préstamos hechos según la hora del día y año.....	73
Figura 36. Configuración del reporte de cantidad de préstamos total de los estudiantes según su sexo	74
Figura 37. Configuración del porcentaje del total	74
Figura 38. Elaboración del reporte de cantidad de préstamos total de los estudiantes según su sexo	75
Figura 39. Configuración del reporte de duración de préstamos según la carrera	75
Figura 40. Elaboración del reporte de duración de préstamos según la carrera	76
Figura 41. Configuración de préstamos de libros según carrera y sexo	77
Figura 42. Elaboración de préstamos de libros según carrera y sexo.....	77
Figura 43. Configuración del reporte de clustering de alumnos según la cantidad de días de préstamo	78
Figura 44. Panel de análisis de Tableau con la opción de clúster	78
Figura 45. Elaboración del reporte de Clustering de alumnos según la cantidad de días de préstamo	79
Figura 46. Código que muestra el top de los 10 libros más recomendados	79
Figura 47. Top 10 de libros más recomendados	80
Figura 48. Línea de tendencia del comportamiento de préstamo de libros por mes	81
Figura 49. Cantidad de préstamos según el mes del año	82
Figura 50. Cantidad de préstamos total según su sexo	83

ÍNDICE DE ANEXOS

Anexo A. Entrevista	93
Anexo B. Presupuesto de la investigación.....	94
Anexo C. Autorización de la investigación	96
Anexo D. Análisis Costo- Beneficio	97
Anexo E. Costo anual de la implementación de la plataforma BI.....	97
Anexo F. Documentación del algoritmo.....	98

RESUMEN

La presente investigación tiene el objetivo de reflejar, según los indicadores dados, la realidad del área, de manera que se tome decisiones estratégicas basada en la data histórica de ella. Los resultados principales fueron los siguientes: Se redujo el tiempo de obtención de los reportes, de 121 minutos a 30 segundos, con lo cual se demostró la reducción del 99.58% del tiempo. Los estudiantes de la escuela de Teología e Ingeniería Civil son quienes más acceden al préstamo de literatura en el Centro de Recursos para el Aprendizaje y la Investigación; el 63.44 % de los préstamos de literatura son a varones mientras el 36.56% son mujeres. Con el comportamiento de resultados y los gustos de los usuarios de biblioteca, el sistema de recomendación (nearest) a través del algoritmo KNN pudo sugerir 10 libros más recomendado para los estudiantes. El desarrollo de la solución se realizó bajo la metodología de Ralph Kimball.

PALABRAS CLAVE: Business Intelligence, sistemas de recomendación, CRAI, toma de decisiones, metodología de Ralph Kimball, inteligencia de negocios, algoritmo de recomendación.

ABSTRACT

The present study aims to reflect, according to the given indicators, the reality of the area, so that strategic decisions can be made based on the historical data. The main results are as follow: The time of obtaining the reports was reduced, from 121 minutes to 5 minutes, which showed a reduction of 96% of the time, the students of the Theology and Civil Engineering school are those who access to the literature the most in the Learning and Research Resource Centre, the 63.44% of the lending of the literature are men while the 36.56% are women. With the behavior of the results and likes of the users of the library, the recommendation system (nearest) through the KNN algorithm was able to suggest 10 more books recommended for students. The development of the solution was made under the methodology of Ralph Kimball.

KEYWORDS: Business Intelligence, recommendation systems, CRAI, decision making, methodology of Ralph Kimball, inteligencia de negocios, recommendation algorithm.

CAPÍTULO I. EL PROBLEMA

1.1. Descripción de la situación problemática

Las autoridades administrativas de la Universidad Peruana Unión, velan por el buen funcionamiento de todos los servicios que ofrece la presente casa de estudios. Uno de los servicios que ofrece a los estudiantes es el acceso al Centro de Recursos para el Aprendizaje y la Investigación (CRAI), el cual es área donde los alumnos de la Universidad Peruana Unión utilizan los recursos bibliográficos como consulta para sus tareas académicas. Las autoridades mencionadas anteriormente, se preocupan porque se fomente la lectura con la finalidad de que los alumnos puedan ser beneficiados con recursos que éste provee.

La lectura de material bibliográfico es de suma importancia para que un estudiante universitario pueda desarrollar capacidades que permitan dar soluciones orientadas al bienestar de la sociedad que le rodea. Sin embargo, la lectura en nuestro país está por debajo del promedio, países como: Chile, Uruguay, Costa Rica, Colombia, México y Brasil nos superan en este aspecto (RPP, 2017).

Según el Anexo A, en la entrevista desarrollada al Jefe de Tecnologías de Información Ing. Walter Luque menciona que en la actualidad el CRAI fomenta la lectura a todas las escuelas profesionales, pero no sabe con exactitud cuáles son las escuelas que más emplean el préstamo de libros, generando un problema, ya que su plan de promover la lectura no está siendo eficiente al no detectar quienes son los que menos acceden a su patrimonio bibliográfico. En los últimos años ha ido guardando copiosas cantidades de data histórica para el manejo de información detallada de los estudiantes y la literatura que usan en su labor académica; para ello se maneja el Sistema Integrado de Biblioteca SirsiDynix Symphony, el cual posee su propia base de datos general que almacena los datos de las tres sedes de la universidad, generando que para sacar los reportes de la data histórica y analizarla, toma bastante tiempo debido al volumen de información que se encuentra en ella, causando la imposibilidad de tomar decisiones de manera rápida. Una necesidad que surge dentro de ésta área, es saber cuáles son las escuelas que menos consumen material bibliográfico, de modo que se pueda solucionar el dilema de saber quiénes son los alumnos pertenecientes a éstas,

son los que raras veces o nunca se prestaron textos de tal forma de que se realice un plan de acción eficiente ante ésta problemática.

Como se ve en el Anexo A, tan sólo el hecho de sacar los reportes en el SYMPHONY demora 5 minutos como máximo (para sacar el reporte de una escuela), y tarda otros 6 minutos como máximo para hacer las consultas en el archivo XML (por escuela). De lo cual podemos deducir que si tuviésemos que sacar el reporte de todas las escuelas profesionales se demoraría 121 min tal como se observa en la siguiente tabla:

Tabla 1

Demora en obtener reportes del sistema

ESCUELA PROFESIONAL	TIEMPO DE DEMORA EN SACAR EL REPORTE EN EL SYMPHONY (Pregunta N°1 Entrevista)	TIEMPO DE HACER LA CONSULTA EN EL ARCHIVO XML (Pregunta N° 4 Entrevista)	TOTAL DE TIEMPO QUE TARDA SACAR REPORTE POR CADA ESCUELA PROFESIONAL
ING DE SISTEMAS	5 min	6 min	11 min
ING CIVIL	5 min	6 min	11 min
ING DE ALIMENTOS	5 min	6 min	11 min
ING AMBIENTAL	5 min	6 min	11 min
CONTABILIDAD	5 min	6 min	11 min
ADMINISTRACIÓN	5 min	6 min	11 min
PSICOLOGÍA	5 min	6 min	11 min
ENFERMERÍA	5 min	6 min	11 min
EDUCACIÓN INICIAL	5 min	6 min	11 min
EDUCACIÓN LINGÜÍSTICA E INGLÉS	5 min	6 min	11 min
EDUCACIÓN PRIMARIA	5 min	6 min	11 min
TOTAL	55 min	66 min	121 min

Fuente: Elaboración propia.

La necesidad de una sincronización entre el universo de los datos históricos y la obtención de variables objetivos ocasiona que los directivos del CRAI no conozcan las tendencias o comportamientos de esa data, al tener este universo de datos se hace complicada la observación, por tal motivo es necesario obtener una muestra de aquel universo, ello también ocasiona que las decisiones tomadas por la gerencia, sea en función a eventos que suceden y no en función a la data histórica que se tiene, de tal manera que es necesario tener una metodología de decisión. Respecto a la transformación de los datos, existe información que ha sido mal ingresada y mal validada, con atributos modelados en la data operacional que no deberían ser considerados, porque no son relevantes a la información que necesita el CRAI, siendo indispensable depurar aquella información.

Esta situación no solamente se presenta en el CRAI, es una problemática que se presenta en la mayoría de empresas peruanas, la toma de decisiones se basa en la intuición y no en la data histórica. Un estudio realizado por la consultora Neo Consulting en el año 2015 sobre 60 empresas líderes arrojó que el 80% de ellas contaban con software de datos, sin embargo, solo el 10% de estas empresas basaba sus decisiones de negocio en estos análisis (PQS PERÚ, 2015) esto indica que solo 5 empresas analizan la información histórica y según a ello toma sus decisiones.

1.2. Justificación

En la actualidad, cuando se habla de Business Intelligence (BI) o Inteligencia de Negocios, se refiere a una terminología muy usada en el ambiente empresarial, la cual aborda aspectos muy importantes que contribuyen en la toma de decisiones dentro de una organización. El CRAI de la Universidad Peruana Unión pertenece al área de servicios y uno de los objetivos específicos de su Plan Operativo Anual es “mejorar la calidad de servicios y atención al usuario” (CRAI, 2016) .La plataforma de BI ayudó a lograr el objetivo mencionado, ya que hoy en día es imprescindible tener una gestión inteligente en un área que desea alcanzar una mayor satisfacción del cliente, en éste caso el cliente es el estudiante universitario de la presente casa de estudios; la implementación de un proyecto de Business Intelligence dentro del CRAI traerá los siguientes beneficios:

- La información que se tenga dentro del área será de calidad, con datos precisos para tomar decisiones acertadas.

- Contribuirá al análisis del comportamiento de los estudiantes hacia el préstamo de los libros.
- Se optimizará el proceso de toma de decisiones.
- Permitirá conocer el estado del área en un periodo concreto del tiempo.
- Se encontrará patrones de similitud entre los libros más prestados, así como los alumnos que acceden a estos préstamos.

Respecto a la ciencia y tecnología la presente investigación conllevará a cumplir otro objetivo específico de su Plan Operativo Anual el cual es: *“Incentivar el uso masivo de los recurso bibliográficos y la investigación por parte de los usuarios motivándolos (CRAI, 2016)”*; al motivar la lectura a quienes raras veces o nunca han accedido al préstamo de algún texto y lograr que ellos empiecen a tener contacto con la literatura, se conseguirá que los alumnos de la Universidad Peruana Unión puedan desarrollar competencias para la innovación e investigación por lo tanto su futura ubicación en el mercado de trabajo será más fácil y en efecto la competitividad de nuestro país se incrementará.

El aporte a la comunidad universitaria será que se mejorará la toma de decisiones dentro del CRAI y por tal efecto se perfeccionará las estrategias de motivar la lectura en los estudiantes. Asimismo, otro aporte que brindará la investigación es que servirá como una guía a seguir implementando plataformas BI en empresas que se dediquen o tengan una biblioteca y estén dispuestas a tener una mejora continua.

En consecuencia, los resultados de la investigación se orientarán a los jefes del área mencionada con el propósito de contar con información detallada sobre el estado y el comportamiento del préstamo de literatura, favoreciendo la optimización del proceso de toma de decisiones en base a la data histórica que nos mostrará la realidad de la organización.

Hoy en día, en el Perú más de 300 empresas utilizan plataformas de BI, por mencionar algunas de las más conocidas tenemos a: Telefónica, Banco del Pacífico, AFP Integra, América TV, Banco del Crédito del Perú (BCP), La Positiva Seguros, Banco Central de la Reserva del Perú, Inkafarma, Grupo Comercio, Universidad Peruana de Ciencias Aplicadas (UPC), Grupo Romero, Presidencia de Consejo de Ministros, Ministerio de Comercio Exterior y Turismo (MINCERTUR), MiFarma, Laive, y muchos otros más que han tenido extraordinarios resultados respecto a la gestión de la información (PREDIQT, 2018).

1.3. Presunción filosófica

La toma de decisiones en base a la data histórica, es algo que Dios nos ha mostrado mediante su santa palabra, la Biblia. Las decisiones tomadas por el pueblo de Israel tenían que basarse en un “escrito está”, es decir, basar sus decisiones en su palabra. La Biblia narra buenas decisiones tomadas por hombres y mujeres obedientes a su palabra y que influyeron en el curso de la historia del pueblo de Israel. Hombres como Abraham, Isaac, Jacob, Moisés, Josué, Gedeón, David mencionados en Hebreos 11 (Valera, 1960) por fe en la palabra de Dios obedecieron y sus decisiones influyen hasta el día de hoy como ejemplo en nuestras vidas.

Así como se toman buenas decisiones hay quienes tomaron decisiones malas basadas en su intuición o en su propio saber y no en la data histórica (La Biblia). En la Biblia se describe como el pueblo de Israel rechazó a Jesús como Mesías, a pesar de que las profecías lo señalaban como el Redentor. En consecuencia, a esa decisión fue que la ciudad de Jerusalén fue sitiada por el general Tito. Elena G. de White comenta en su libro Conflicto de los siglos “Ni un solo cristiano pereció en la destrucción de Jerusalén. Cristo había prevenido a sus discípulos, y todos los que creyeron sus palabras esperaron atentamente las señales prometidas. “Cuando viereis a Jerusalén cercada de ejércitos —había dicho Jesús, sabed entonces que su destrucción ha llegado. Entonces los que estuvieren en Judea, huyan a los montes; y los que en medio de ella, váyanse (S. Lucas 21: 20, 21.)” (White, 1911).

Como cristianos, es necesario tomar nuestras decisiones en la data histórica mostrada en la palabra de Dios ya que es la única salvaguardia que nos ayudará en nuestra vida espiritual contra las acechanzas del enemigo.

1.4. Objetivos de la investigación

1.4.1. Objetivo general

Implementar una plataforma de Business Intelligence para la toma de decisiones en el Centro de Recursos para el Aprendizaje y la Investigación (CRAI) de la Universidad Peruana Unión.

1.4.2. Objetivos específicos:

1. Elaborar el proyecto de solución de la plataforma BI.
2. Elaborar el pre tratamiento de los datos, generando el dataset a ser analizado.
3. Implementar una plataforma de explotación de la información.

CAPÍTULO II. BASES TEÓRICAS

2.1. Revisión de la literatura

2.1.1. Metodología de Ralph Kimball vs Bill Inmon

A la hora de establecer los pasos a seguir para construir una solución, existen desacuerdos entre estas dos metodologías, sus principios son muy diferentes en su estructura interna y el alcance así como su intención y finalidad (Dertiano, 2016).

Por ello se analiza los siguientes aspectos:

- Presupuesto para desarrollar el proyecto.
- Plazos adecuados
- Alcance.
- Dificultad de las labores de mantenimiento.

La siguiente tabla expone cómo estos factores afectan a los dos modelos de data warehouse.

Tabla 2

Características de las metodologías

Metodología	Características
Ralph Kimball	Coste inicial bajo, tiempo de desarrollo inferior, necesita equipo con especialización media, el alcance es a departamentos individuales, de mantenimiento más complejo.
Bill Inmon	Coste inicial alto, requiere más tiempo de desarrollo, necesita equipo con especialización alta, su alcance es a toda la compañía, de fácil mantenimiento.

Fuente: Elaboración propia.

Como se ve con anterioridad, la construcción de una solución con Ralph Kimball está enfocado a la consulta de la información, por lo que está estructurado para garantizar una explotación de datos rápida y sencilla, no prescindiendo de personas especializadas para el mismo. Por otro lado, Bill Inmon pretende la integración de todos los datos de la compañía con el almacenaje de grandes volúmenes de datos por lo que está estructurado para evitar la redundancia de datos y simplificar el mantenimiento de ellos. De todo lo presentado, podemos concluir que la metodología de Kimball se ajusta más a proyectos pequeños y la metodología de Inmon es más apropiado para proyectos de mayores dimensiones(Dertiano, 2016). Según la *Tabla 2* la investigación presente concluye el uso de la metodología Ralph Kimball debido a:

- Se ajusta más al proyecto de investigación al poder ser implementado en proyectos de menor escala.
- El tiempo de desarrollo del proyecto es inferior, así como el coste inicial de éste es bajo.

2.1.2. Herramientas a utilizar:

2.1.2.1. Pentaho Data Integration (Kettle)

Pentaho Data Integration, la solución tecnológica conocida como Kettle, es una de las herramientas ETL de código abierto más poderosas y versátiles cuando se diseñan procesos de integración adaptados a las necesidades de la empresa, principalmente con el objetivo de construir y explotar su Data Warehouse (DW), entre otros.

Uno de sus aspectos más ventajosos consiste en su capacidad de adaptación, que permite que los procesos de integración de datos se desarrollen fácilmente en el datawarehouse, el centro de la arquitectura de los sistemas de información, entendido como el proceso que consiste en su recopilación, cruce y enriquecimiento; ofrece al usuario la extracción, transformación y carga de datos; todo ello de manera automática evitando el trabajo manual (Durán, 2017). Tiene un diseño intuitivo con una arquitectura escalable basada en estándares (Goncalves, 2017).

Como herramienta de ETL, Kettle es visual y logra una mayor productividad del equipo de desarrollo y mantenimiento, lo que le permite conectarse a diferentes fuentes y arquitecturas para cargar con precisión los datos en un repositorio, aunque también se puede

usar para muchas otras funciones, como sincronizar, enmascarar, migrar datos entre diferentes aplicaciones o, por ejemplo, actualizar a los usuarios a sistemas paralelos.

A continuación, se describen las ventajas de la herramienta:

- Es open source: su uso no implica un gasto en licencias de software, lo que puede ser un punto a favor, especialmente para las empresas nuevas.
- Edición de pago: También hay una edición de pago (Pentaho BI Suite Enterprise Edition) que incluye servicio técnico y brinda soluciones más completas que el software gratuito.
- Amplia comunidad: Tiene actividad intensa a nivel de pruebas beta como en foros, muy útil para la consulta de posibles problemas, aunque obtener el máximo rendimiento implica un mínimo de especialización.
- Tecnologías innovadoras y fáciles de usar: Tanto en la versión de pago como en la gratuita, se trabaja con las últimas tecnologías (creadas en las plataformas de desarrollo Java y MySQL), sin descuidar una instalación y configuración simples.
- Múltiples programas: Ser modular y tener aplicaciones para realizar funciones muy diferentes, incluida la integración de datos, es una de sus grandes ventajas, que permite el uso conjunto e independiente en diferentes plataformas.

Es, en resumen, Pentaho Data Integration (Kettle) es una opción que puede ser interesante a la hora de elegir el software para realizar los procesos de integración de una base de datos, con el objetivo final de la integración inteligente de datos para ayudarnos a lograr la calidad correcta, capaz de maximizar su valor de manera operativa y global.

2.1.2.2. TalendETL

Talend es una plataforma de integración de datos de código abierto. Proporciona diversos software y servicios para la integración de datos, gestión de datos, integración de aplicaciones empresariales, calidad de datos, almacenamiento en la nube y Big Data. Talend entró en el mercado por primera vez en 2005 como el primer proveedor comercial de software de código abierto de software de integración de datos. En octubre de 2006, Talend lanzó su primer producto: Talend Open Studio, que actualmente se conoce como Talend

Open Studio para la integración de datos. Desde entonces, ha lanzado una amplia gama de productos que se utilizan favorablemente en el mercado. Talend es considerado como el líder de la próxima generación en la nube y el software de integración de Big Data. Ayuda a las empresas a tomar decisiones en tiempo real y a ser más orientadas a los datos. Con Talend, los datos se vuelven más accesibles, su calidad mejora y puede trasladarse rápidamente a los sistemas de destino (Chand, 2019). A continuación, se muestra algunas ventajas éste:

- Desarrollo e implementación más rápidos: Talend automatiza las tareas y las mantiene para usted.
- Menos gastos: Talend proporciona herramientas de código abierto que se pueden descargar de forma gratuita. Además, a medida que los procesos se aceleran, los costos de desarrollo se reducen considerablemente.
- Plataforma unificada: Talend satisface todas nuestras necesidades bajo una base común para los productos basados en las necesidades de la organización.
- Gran comunidad: al ser un software de código abierto, está respaldado por una gran comunidad. Es el lugar preferido para todos los usuarios de Talend y miembros de la comunidad donde pueden compartir información, experiencias, dudas, consultas, etc.

En la presente investigación se utilizará diversas herramientas tecnológicas, tanto para la transformación y limpieza de los datos, así como para los reportes de los resultados encontrados. En la siguiente tabla se muestra las herramientas para la ejecución del ETL y en la *Tabla 3* se muestra las herramientas para la visualización de los reportes.

Tabla 3

Herramientas para la transformación y limpieza de datos

Herramientas	Características
Kettle	<p>La herramienta es muy intuitiva y con unos conceptos básicos se puede empezar a ejecutar un proyecto.</p> <p>Es un motor de transformación, ha sido diseñado con la finalidad de cubrir la necesidad de la integración de datos.</p> <p>Respecto a su performance es más rápida ya que no tiene que estar moviendo código generado.</p> <p>Respecto al manejo de errores, es capaz de solucionar el problema y continuar de donde fue dejado el trabajo.</p>
TalendETL	<p>La herramienta es poco intuitiva y difícil de entender, pero una vez superada la curva de aprendizaje se puede entender las grandes posibilidades y la potencia de ésta.</p> <p>Es un generador de código y ello hace que se dependa del lenguaje elegido en los proyectos (Java o Groovy) necesitando un nivel alto de éste para tener la mayor potencia de ésta herramienta.</p> <p>Respecto a su performance es más lento ya que se demora en la carga de código Java generado.</p> <p>Respecto al manejo de errores, si el trabajo falla debe solucionarse el problema y reiniciarse el proyecto.</p>

Fuente: Elaboración propia

Según la Tabla 3 , el investigador ha visto conveniente la herramienta Kettle para la transformación y limpieza de datos sobre TalendETL a causa de que Kettle es muy intuitiva, de mejor performance y manejo de errores.

2.1.2.3. Tableau

Es una plataforma de análisis integral de Business Intelligence que permite convertir los datos en información de utilidad para la empresa (Tableau Software, 2017) Facilita la integración de diversos tipos de datos, transformándolos en dashboards o tableros que favorecen la toma de decisiones a partir de la información generada, también provee la opción de integrar datos de Google Analytics o de Excel (Fernández, 2018).

Gartner, una consultora norteamericana de investigación de tecnologías, elabora el cuadrante mágico de Gartner; el cual es un gráfico donde se ubican las principales empresas proveedoras de servicios tecnológicos del mundo, el objetivo principal del cuadro es mostrar los productos impulsan la mayoría de las compras en el mercado actual que cumplen con los

criterios de una plataforma de analytics y BI. Tableau lidera dicho cuadrante y este reconocimiento ha sido dado por las opiniones de los clientes (Microsystem, 2018). Dicho cuadrante lo podemos observar en la *Figura 1*

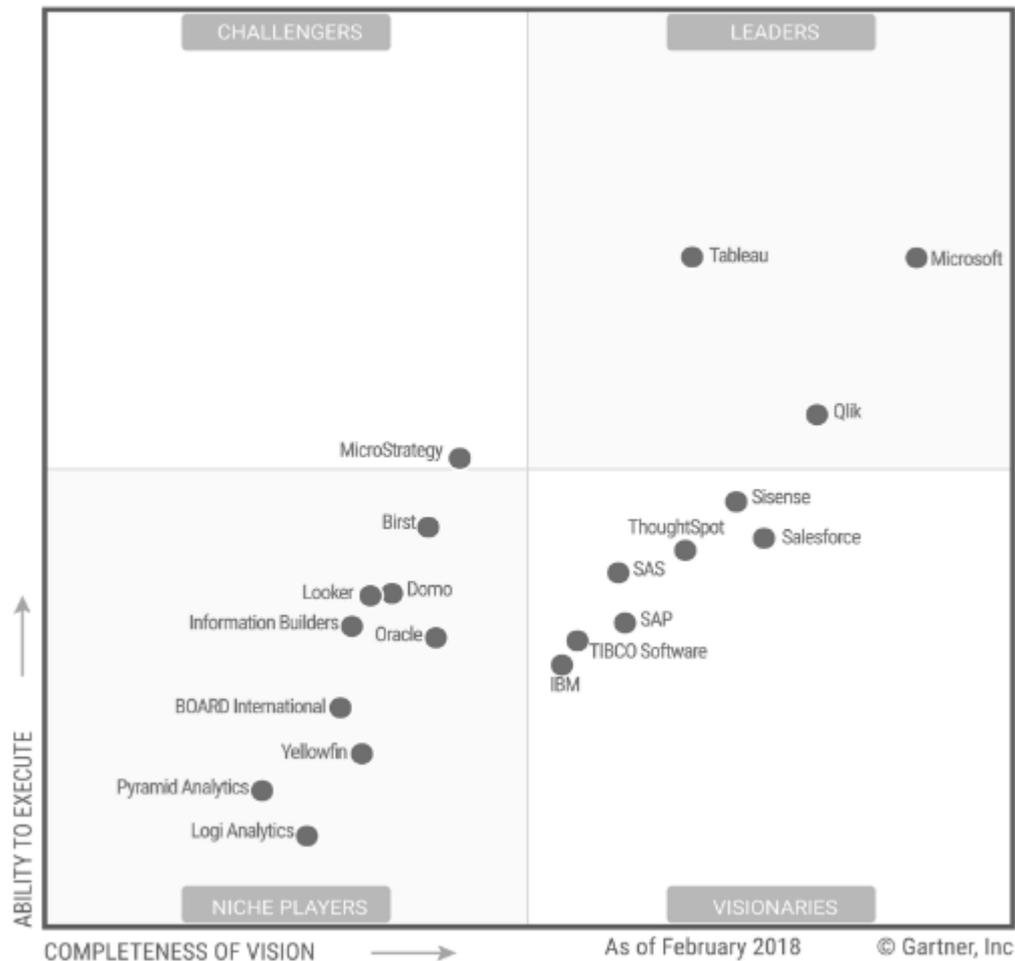


Figura 1. Cuadrante mágico para plataformas de análisis e inteligencia de negocios. Fuente: (Gartner, 2018).

¿Qué hace el software de visualización Tableau y por qué es tan específico que muchas empresas lo incorporan a sus operaciones diarias? Una poderosa herramienta de inteligencia empresarial para la visualización de análisis, presenta los resultados en numerosas formas vívidas para obtener una mejor visión. Tableau tiene versiones de escritorio y en línea, que permiten el acceso a los datos de software en la nube o en las instalaciones. La funcionalidad de la solución no está limitada por la representación de información gráfica; También hay un montón de trabajo bajo la superficie. De hecho, la aplicación realiza solicitudes a la nube y bases de datos relacionales, hojas de cálculo y cubos

OLAP para los datos estadísticos requeridos, y luego los analiza, clasifica y correlaciona para generar un informe analítico completo.

Tableau ofrece las siguientes bondades:

- **Notable capacidad de visualización:** Por supuesto, las capacidades incomparables de visualización de información se encuentran en la parte superior de la lista de beneficios del software Tableau. La calidad de visualización de datos de la aplicación es superior a la que ofrecen los competidores del software Tableau. Incluso los productos de los proveedores tradicionales de inteligencia empresarial como Oracle Data Visualization o los productos de IBM para la representación de datos, no pueden competir con la calidad de diseño e ilustración que proporciona Tableau. Convierte la información estadística no estructurada en resultados lógicos integrales, que son cuadros de mandos totalmente funcionales, interactivos y atractivos. Están disponibles en varios tipos de gráficos y son fáciles de usar en negocios
- **Facilidad de uso:** La manera intuitiva de la creación de gráficos y una interfaz fácil de usar permiten a los usuarios que no son desarrolladores utilizar la funcionalidad básica de la aplicación al máximo. Los usuarios organizan los datos sin procesar en diagramas pegajosos de una forma que permite arrastrar y soltar, lo que facilita el análisis de la información y elimina la necesidad de la ayuda de un departamento de TI para la creación de patrones. Los usuarios no profesionales pueden disfrutar de las capacidades que ofrece Tableau para el análisis de estadísticas, como el desarrollo de paneles de control, etc., sin una capacitación profunda. Sin embargo, para entrar en las capacidades de la solución, un conocimiento más profundo es una necesidad. Además, la estrecha participación de los especialistas en TI es una necesidad si una empresa busca expandir la funcionalidad de la solución.
- **Alto rendimiento:** Además de su alta funcionalidad de visualización, los usuarios califican su rendimiento general como robusto y confiable. La herramienta también funciona rápidamente incluso con Big Data, lo que hace que su potente rendimiento sea un punto importante en la lista de las ventajas de Tableau.
- **Conexiones de múltiples fuentes de datos:** El software permite establecer conexiones con muchos orígenes de datos, como HADOOP, SAP y DB Technologies, lo que mejora la calidad del análisis de datos y permite la creación de un panel informativo

unificado. Dicho panel de control otorga acceso a la información requerida para cualquier usuario.

- Comunidad y foros: La cantidad de seguidores de Tableau que invierten su experiencia y habilidades en la comunidad aumenta constantemente. Los usuarios de negocios pueden reforzar sus conocimientos sobre análisis de datos e informes y obtener muchos conocimientos útiles en esta comunidad. Además, los visitantes del foro están listos para ayudar a resolver cualquier problema de los usuarios y compartir su experiencia.
- Aplicación móvil: Y la última en nuestra lista de beneficios principales de Tableau, existe una aplicación móvil eficiente disponible para iOS y Android. Agrega movilidad a los usuarios de Tableau y les permite mantener las estadísticas a su alcance, así como también admite la funcionalidad completa que tienen las versiones de escritorio y en línea.

Tableau ofrece diversas plataformas tales como: Tableau Desktop y Tableau Server. Tableau Server le permite administrar de forma centralizada la colaboración, la seguridad y el rendimiento. Dos escenarios principales desencadenan la necesidad de Tableau Server:

- Tiene un entorno incontrolado de alto crecimiento y desea más seguridad, colaboración, gobierno y rendimiento.
- Desea dominar la cantidad de fuentes de datos que se utilizan junto con la proliferación de paneles y libros de trabajo, cada uno con sus propias capas de metadatos, extractos y copias.

Estas dos plataformas se diferencian en que Tableau Desktop está instalado en una estación de trabajo o computadora portátil. Tableau Server, por otro lado, está instalado en un servidor Windows y es accesible para los usuarios a través de un navegador. A continuación, se enumeran las capacidades en Tableau Server que no están en Tableau Desktop y la funcionalidad que está presente en Tableau Desktop, pero no en Tableau Server.

- Tableau Desktop tiene el entorno de desarrollo más completo y permite la creación de fuentes de datos. Tableau Server tiene un modo de creación web similar a Tableau Desktop y también proporciona funciones adicionales de colaboración, seguridad y escalabilidad que no están disponibles en Tableau Desktop.

Las diferencias entre los productos caen dentro de estas categorías:

- Colaboración: En el entorno de creación web 10.x de Tableau Server, los usuarios pueden crear, editar y publicar libros de trabajo sin la necesidad de Tableau Desktop. Los libros de trabajo publicados pueden ser compartidos y modificados por usuarios autorizados. Esta funcionalidad permite a las organizaciones acercarse al Santo Grial del análisis de autoservicio. Con Tableau Server, las fuentes de datos, los libros de trabajo y los paneles de control se pueden consolidar en un solo proyecto y pueden administrarse de manera centralizada, eliminando el procedimiento propenso a errores de enviar libros de trabajo y enlaces por correo electrónico.
- Fuentes de datos centralizadas: Las fuentes de datos y los metadatos solo se pueden crear en Tableau Desktop. Estas fuentes de datos se pueden publicar en Tableau Server o Tableau Online, creando así una "fuente única de la verdad". "En Tableau Server, una vez que se validan las fuentes de datos, se pueden "certificar". Tableau Desktop debe usarse para la mayoría de las modificaciones a las fuentes de datos.
- Aplicaciones móviles: En Tableau Desktop, los usuarios pueden crear diseños específicos para dispositivos para iPhone, iPad y Android. La creación de diseños específicos del dispositivo para Dashboards solo está disponible en Tableau Desktop. Cuando el panel de control se publica en Tableau Server, los usuarios pueden ver, interactuar y suscribirse al panel de control, y se representará de manera diferente para cada dispositivo específico. Si desea crear, ver o editar contenido de Tableau en un iPad, necesitará Tableau Server y Tableau Mobile. Tableau Desktop y Tableau Reader no están disponibles en iPad.
- Seguridad: Tableau Server se puede usar para administrar el acceso al contenido publicado para ayudar a proteger los datos confidenciales. Los administradores pueden establecer grupos y permisos de usuario en proyectos, libros de trabajo, vistas y fuentes de datos. Si necesita restringir los datos mostrados en Tableau por usuario, hay varias formas de lograrlo. Algunos métodos requieren Tableau Server y otros solo funcionan en Tableau Desktop; para obtener más detalles, consulte nuestro blog [Asegurar y filtrar datos por usuario en Tableau](#).

- Funcionalidad en el entorno de creación: Tableau Server y Tableau Desktop ofrecen un rico entorno de desarrollo / creación. Aunque son muy similares en términos de funcionalidad, aún no es posible realizar todo su trabajo de desarrollo en Tableau Server.

A continuación, se muestra una lista de las principales diferencias entre Tableau Server 10.xy Tableau Desktop desde una perspectiva de desarrollo:

- a) En Tableau Server, el administrador debe otorgar permisos de edición web a usuarios o grupos para habilitar la creación web. Una vez que los usuarios tienen los permisos, pueden crear nuevos libros a partir de fuentes de datos publicados o modificar los libros existentes.
- b) Las nuevas fuentes de datos solo se pueden publicar desde Tableau Desktop.
- c) La edición de la fuente de datos está limitada en Tableau Server. Solo se puede editar la información de "conexión".
- d) Tableau Desktop proporciona capacidades analíticas que no se encuentran en el entorno de creación web de Tableau Server. Por ejemplo, las características de pronóstico y agrupación faltan en el panel de análisis de Tableau Server. Además, los usuarios no pueden editar sus líneas de tendencia en Tableau Server.
- e) En Tableau Server no puede hacer doble clic en el título de una hoja de trabajo para editarlo.
- f) La información sobre herramientas y la edición y el formato están limitados en Tableau Server. Además, la creación de "Voz en una información sobre herramientas" de Tableau 10.5 solo está disponible en Tableau Desktop.
- g) Las acciones del panel solo se pueden crear en Tableau Desktop.
- h) No puede especificar el formato en el nivel del panel de control o en el nivel de la hoja de trabajo en Tableau Server.
- Escalabilidad / rendimiento: Tableau Server proporciona licencias basadas en el núcleo y la capacidad de agregar fácilmente más servidores para manejar la carga de usuarios. Las vistas de licencia proporcionan visibilidad del uso y le permiten comprender las necesidades analíticas de sus usuarios, ajustando la combinación para reducir los costos y aumentar la adopción de los usuarios. No puede obtener esta visibilidad de uso y administración sin Tableau Server.

2.1.2.4. Google Data Studio

Es una herramienta gratuita de Google que ayuda a la visualización de datos mediante cuadros de mando, la cual nos permite la consulta de datos de diversas fuentes extrayendo información útil, además da la opción de configurar los gráficos de los reportes generados (Google, 2018). Brinda todo lo que necesita para convertir los datos analíticos de su cliente en informes informativos y fáciles de entender a través de la visualización de datos. Una de las cosas interesantes de Data Studio es que puede obtener más que solo datos de Google Analytics, incluso puede importar datos de Facebook, siempre que incluya esa información en una hoja de Google. Los informes también son dinámicos, por lo que cuando hay una actualización del origen de datos, la información actualizada / nueva aparece automáticamente en cualquier informe que haga referencia al origen. Además, los informes se pueden compartir, por lo que puede otorgar permiso a las personas para ver los informes y / o permitirles hacer cambios. Google Data Studio utiliza la misma funcionalidad que Google Docs y Google Sheets, por lo que todo lo que necesita hacer es presionar el botón "Compartir" para que sus clientes u otros miembros de su equipo puedan ver o editar los informes.

A continuación, se describe las ventajas principales de esta herramienta:

- **Personalizable y eficiente:** Antes de la disponibilidad y la adopción generalizada de Data Studio, muchos profesionales de marketing confiaban en la creación de informes a través del panel de control estándar de Google Analytics. Esta plataforma requería que los usuarios hicieran clic entre muchas pantallas y seleccionaran entre muchos filtros para crear informes. Un proceso que consume tiempo. Sin embargo, para poder asignar los presupuestos de manera eficiente, los tomadores de decisiones requieren un cambio rápido en grandes cantidades de datos precisos. Los informes deben ser producidos rápidamente. También deben ser fáciles de leer y comprender. Data Studio se basa en estos conceptos básicos, por lo que es nuestra opción preferida para el informe de datos.
- **Fuentes de datos múltiples:** Una de las características más importantes y útiles de Data Studio es su capacidad para extraer información de una variedad de fuentes. Algunas de las fuentes de datos más populares incluyen AdWords, Analytics, YouTube, Search Console, BigQuery y MySQL.

- Colaboración: Data Studio funciona como cualquier otro documento en Google Drive. El autor puede otorgar acceso a compañeros con la opción de permitir o restringir la edición. También hay una opción para guardar y compartir contenido como un archivo PDF.
- Filtrado y actualizaciones en tiempo real: El panel de control de Data Studio se actualiza en tiempo real. Los usuarios pueden elegir cambiar el intervalo de fechas con el clic de un botón. Además de esto, los usuarios pueden seleccionar entre las opciones avanzadas de filtrado que permiten una vista rápida de los conjuntos de datos relevantes (Braga, 2018).
- Gratuito: Google Data Studio es completamente gratuito. Google está haciendo un trabajo fantástico de construir una gran cantidad de herramientas intuitivas y que ahorran tiempo. Cuando se lanzó Data Studio por primera vez, había un límite de 5 informes para los usuarios. Este gorro hace tiempo que se ha eliminado. Ahora, cualquiera puede hacer uso de estos cuadros de mando competentes sin ninguna limitación. Gratis. En general, Data Studio ayuda a los usuarios a crear y monitorear informes, algo que ayuda a hacer la vida más fácil en el lugar de trabajo.

2.1.2.5. Qlikview

Es una plataforma de Business Intelligence de autoservicio que ofrece el análisis de datos y con ello se facilita la toma de decisiones dentro de una empresa u organización de forma colaborativa, está diseñado para crecer al ritmo del negocio mostrando escalabilidad (SAIMA Solutions, 2017).

A continuación, se muestran las ventajas de esta herramienta:

- Interpretación y análisis de datos: Sus paneles y aplicaciones visuales y dinámicas ayudan en la interpretación y el análisis de datos de forma rápida y eficiente.
- Compartir datos: Diferentes equipos pueden colaborar para tomar decisiones cruciales sobre información e interpretaciones basadas en datos. Esta colaboración se ve facilitada por el análisis social y la capacidad de compartir datos en tiempo real de Qlikview.
- Bajo mantenimiento: La implementación del software es muy simple y sin complicaciones. Incluso en los últimos años, el software no necesita mucho mantenimiento o supervisión técnica.

2.1.2.6. *MicroStrategy*

Es una empresa que brinda software OLAP de Business Intelligence, permite el análisis y visualización de los datos mediante cuadros de mando, el aprendizaje automático mediante una aplicación así como la localización inteligente mediante mapas (MicroStrategy, 2018).

A continuación, se detallan los beneficios que ofrece esta herramienta:

- **Funciones múltiples:** combina la capacidad de generar informes y monitorear datos en tiempo real y la capacidad de manipular bases de datos en un paquete. El software puede generar cuadros de mando, cuadros de mandos, notificaciones e informes empresariales. La amplia gama de funciones es beneficiosa para los usuarios, ya que no requiere un conocimiento avanzado de la programación de bases de datos u otras habilidades complejas, ya que el software permite a los usuarios simplemente apuntar y hacer clic para acceder a las funciones. Las funciones de fácil acceso hacen que trabajar con diferentes tipos de información sea más conveniente.
- **MicroStrategy Web:** La mayor ventaja de la versión web del software es la capacidad de los usuarios para acceder a la información comercial a través de la aplicación en línea, según MicroStrategy. Los usuarios pueden ordenar la información con una barra de navegación para moverse de una carpeta a otra. El programa incluye un cuadro para ingresar palabras clave, lo que facilita la búsqueda de una gran cantidad de datos y es similar a los cuadros de búsqueda de los navegadores web. El diseñador de informes permite a los usuarios crear informes personalizados de información crucial.
- **MicroStrategy Mobile:** El software móvil de MicroStrategy ofrece a sus usuarios el beneficio de la compatibilidad con muchos dispositivos móviles conocidos. Los usuarios de MicroStrategy Mobile pueden crear informes en estos dispositivos móviles. El software es compatible con teléfonos iPhones de Apple y iPad, de acuerdo con la Guía de ECRM. Los usuarios de Apple pueden descargar el programa como una aplicación gratuita en la tienda iTunes Store y comenzar a aprender sobre el programa con los tutoriales incorporados.

De todas las herramientas anteriormente presentadas, podemos resumir sus características en la siguiente tabla:

Tabla 4

Herramientas para la visualización de los reportes

Herramienta	Intuitiva	Demora la curva de aprendizaje	Es gratuito
Google Data Studio	SI	NO	SI
Tableau	SI	NO	SI
Qlikview	SI	SI	NO
MicroStrategy	SI	NO	NO

Fuente: Elaboración propia

Según la *Tabla 4* el investigador ha visto oportuno elegir dos herramientas para la visualización de los reportes: Tableau y Google Data Studio, siendo que por sobre las demás estas presentan una interfaz intuitiva y amigable, tienen la capacidad de ejecutar cálculos complejos, así como el análisis integrado avanzado de los datos. Otra de las razones es que las herramientas seleccionadas nos permiten acceder a ella de manera libre de pago, por ejemplo, podemos obtener Google Data Studio con solo una cuenta de correo de Google y respecto a Tableau otorga una licencia a los estudiantes universitarios para acceder a su herramienta.

Anselmo & Espinoza (2015) en su investigación “Implementación de un Datamart para toma de decisiones de recibos distribuidos a clientes pospago de Telefónica del Perú” usa como herramienta de solución BI a Tableau así como Garcés (2015) en su investigación “Estudio comparativo de metodologías e implementación de alternativas business intelligence opensource vs. propietarias en entornos tradicionales; caso prototipo en las pymes en el sector agroindustrial”.

2.1.2.7. Microsoft Excel

Es un programa incorporado al paquete de Microsoft Office, con funciones de hojas de cálculo, en el cual se puede crear tablas, calcular y analizar datos, su funcionamiento es en el sistema operativo Windows (OBS, 2018). Se pueden almacenar base de datos, administrarlos y presentarlos en diferentes formatos (Excelfull, 2018).

2.1.2.8. *PyData*

Es un programa educativo de NumFOCUS, la cual es una organización sin fines de lucro, éste provee un foro para la comunidad internacional de usuarios y desarrolladores de herramientas de análisis de datos; promueve la discusión de buenas prácticas para la gestión, procesamiento, análisis y visualización de datos (PyData, 2016). Una de las suites gratuitas que nos ofrece es ANACONDA, la cual es una distribución de Python que contempla diferentes aplicaciones y librerías para el desarrollo de ciencia de datos con Python, ésta contiene más de 1000 paquetes de datos (Watkins, 2018).

2.1.2.9. *Spyder*

Es uno de los mejores IDEs de Python que se puede usar en ciencia de datos. Es ligero y capaz de ejecutar secuencias de comandos complejas de Python, en términos de rendimiento informáticos (González, 2018). A continuación, se presenta las características y ventajas de esta herramienta:

Tabla 5

Características y ventajas de Spyder

Características	Ventajas
IDE simple y liviano de fácil instalación, con documentación detallada	Interfaz en español e inglés
Editor de código abierto, permite resaltado de sintaxis.	Se encuentra todo en un solo lugar en un ambiente sencillo y fácil de usar.
Compatible con la depuración en tiempo de ejecución, esto quiere decir que los errores se ven en pantalla tan pronto como se los escribe.	Creador por y para científicos.
Se integra con alguna de las librerías fundamentales, tales como, NumPy, Matplotlib, SciPy, entre otras.	Es una buena puerta de entrada para aprender Python. Para estudiantes de pregrado y colegas de trabajo.

Fuente. Basado en Córdoba (2015) & González (2018).

2.1.2.1. *Dataframe*

Un dataframe se interpreta como una hoja de datos o un marco de datos en los que cada fila corresponde a un sujeto y cada columna a una variable, el cual es utilizado normalmente cuando se realiza un estudio estadístico sobre los sujetos de una muestra. Su estructura es similar a la de una matriz (Santana & Nieves, 2014), también es similar a las tablas de bases de datos relacionales como SQL (Moya, 2015).

2.1.2.2. Extracción, Transformación y Carga

ETL es la abreviatura en inglés de: Extracción, Transformación y Carga. Permite extraer datos del entorno origen, transformarlos según nuestras necesidades de negocio para integración de datos y cargar estos datos en los entornos destino. Los entornos origen y destino son usualmente bases de datos y/o ficheros, pero en ocasiones también pueden ser colas de mensajes de un determinado middleware, así como ficheros u otras fuentes estructuradas, semiestructuradas o no estructurada basada en técnicas de consolidación (Curto, 2011).

En sus funcionalidades básicas de ésta herramienta, está:

- Gestión y administración de servicios.
- Extracción de datos.
- Transformación de datos.
- Carga de datos.
- Gestión de datos.

En la figura 2 se muestra el esquema del proceso ETL.

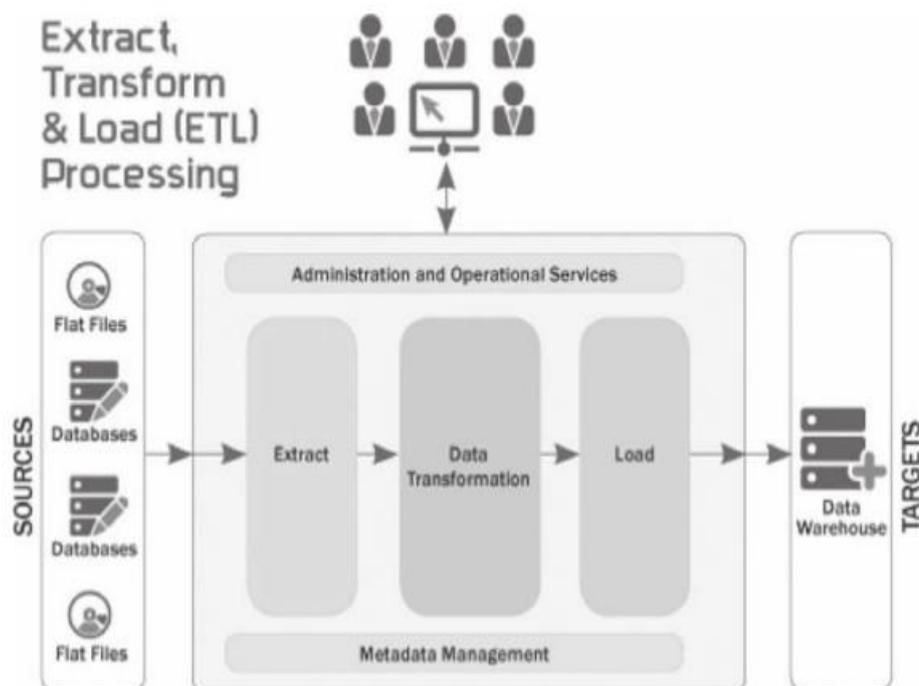


Figura 2. Esquema del proceso ETL.

Fuente:(Curto, 2011).

2.1.2.3. Machine learning

Es una rama de la inteligencia artificial, responsable de crear programas de software capaces de aprender de manera automatizada, puede identificar patrones complejos en grandes cantidades de datos, también permite predecir comportamientos a través de algoritmos que tienen la capacidad de optimizarse de manera independiente (Castro, 2017). En la presente investigación se utilizará el aprendizaje supervisado, el cual es uno de los tipos de machine learning.

2.1.2.4. Sistemas de recomendación

Un sistema de recomendación, es un sistema inteligente, capaz de aprender las preferencias o gustos de un usuario y poder ofrecerle aquella información que pueda ser de utilidad para dicho usuario. Por tanto, se entiende el sistema de recomendación como un filtro que deja pasar aquella información que le va a resultar de interés a los usuarios y va a desechar aquella información que le pueda resultar al usuario indiferente (Moya, 2013). Uno de los tipos de sistema de recomendación que se utilizará es el de filtrado colaborativo.

2.1.2.5. Filtrado colaborativo

El sistema de recomendación de tipo filtrado colaborativo consiste en visualizar que usuarios son similares al usuario activo y luego recomendar aquellos ítems que no han sido valorados por el usuario activo y que han sido bien votados por usuarios similares (Moya, 2013). A continuación, se mostrará dos métricas para determinar la similaridad entre los usuarios, calculan las distancias entre los mismos:

1. Diferencia cuadrática media (MSD)

Permite el cálculo del error del modelo, de cuan preciso es el sistema colaborativo

Ecuación 1 Diferencia cuadrática media

$$sim(x, y) = 1 - \frac{1}{\#B_{x,y}} \sum_{i \in I_u} \left(\frac{r_{x,i} - r_{y,i}}{max - min} \right)^2 \in [0,1]$$

Siendo $#B_{x,y}$ el número de el número de libros que ha sido calificados (y que tiene que ser necesariamente mayor que 0); siendo $r_{x,i}$ y $r_{y,i}$ los votos x e y respectivamente, y siendo \max y \min las el puntaje máximo y mínimo según calificación dada.

2. *Distancia coseno*

Mide la distancia que hay entre dos usuarios en función del ángulo que forman entre ellos (Moya, s. f.).

3. *Algoritmo KNN*

Es un algoritmo de aprendizaje supervisado, es decir, que a partir de un juego de datos inicial su objetivo será el de clasificar correctamente todas las instancias nuevas. El juego de datos típico de este tipo de algoritmos está formado por varios atributos descriptivos y un solo atributo objetivo (también llamado clase) (Ad Tech & Analytics, 2017).

2.1.2.6. *SirsiDynix Symphony*

Es un sistema integral de gestión para bibliotecas robusto, tiene todas las funcionalidades y herramientas que necesita una biblioteca moderna de cualquier magnitud, permite las adquisiciones y catalogación de los libros. Éste cuenta con una plataforma web, como es el caso de Symphony Web que no depende de instalación o actualizaciones. Además posee su versión escritorio tal como es Workflows (SirsiDynix, 2017).

2.1.2.7. *Catalogación de los libros del CRAI*

1. *Sistema de clasificación decimal Dewey*

Melvil Dewey fue un bibliotecario que en el año 1876 creó un sistema numérico decimal para la organización de los libros de la biblioteca escolar donde laboraba, Dewey dividió el conocimiento en diez grandes categorías tal como lo muestra la siguiente tabla. Cada una de las cifras se puede subdividir varias veces para lograr identificar el tema del libro, éste sistema permite la distribución de los libros en los estantes, de manera que toda la literatura que traten de un tema específico esté ubicado en un mismo lugar (Biblioteca Pública de Denver, s. f.).

Tabla 6

Divisiones principales del sistema de clasificación decimal Dewey.

Número	Clasificación	Tema que incluye
000	Obras generales	Enciclopedias, periodismo, bibliotecas, museos.
100	Filosofía y psicología	Sentimientos, emociones, valores
200	Religión	Biblia, cristianismo y otras religiones, mitos.
300	Ciencias sociales	Economía, educación, adivinanzas, leyendas, días feriados.
400	Lenguaje	Diccionarios, lenguajes de señas.
500	Ciencia	Matemáticas, plantas, animales, volcanes, huracanes, experimentos, imanes, electricidad.
600	Tecnología y ciencias de la salud	Cuerpo humano, medicina, enfermería, inventos, agricultura, crianza de mascotas, trenes, aviones, cocina.
700	Arte y recreación	Pintura, fotografía, artesanías, manualidades, dibujo, música, deportes, juegos.
800	Literatura	Poesía, cuentos, novelas, teatro.
900	Historia y geografía	Atlas, mapas, banderas, castillos, biografías, países, grupos étnicos.

Fuente: (EcuRed, s. f.).

2. Tabla de Málaga

La creación de la tabla de Málaga ésta surgió de la necesidad de clasificar campos como: religión e iglesia, derecho romano, literatura, historia de la geografía del Perú y América Latina; ya que estos campos no eran contemplados en el Sistema de clasificación decimal Dewey porque éste es de origen anglosajón, tampoco abarcaba los apellidos de origen hispano. Por ello Luis F. Málaga creó esta tabla que permite que las bibliotecas hispanas puedan clasificar sus libros (Mundobiblio, 2017).

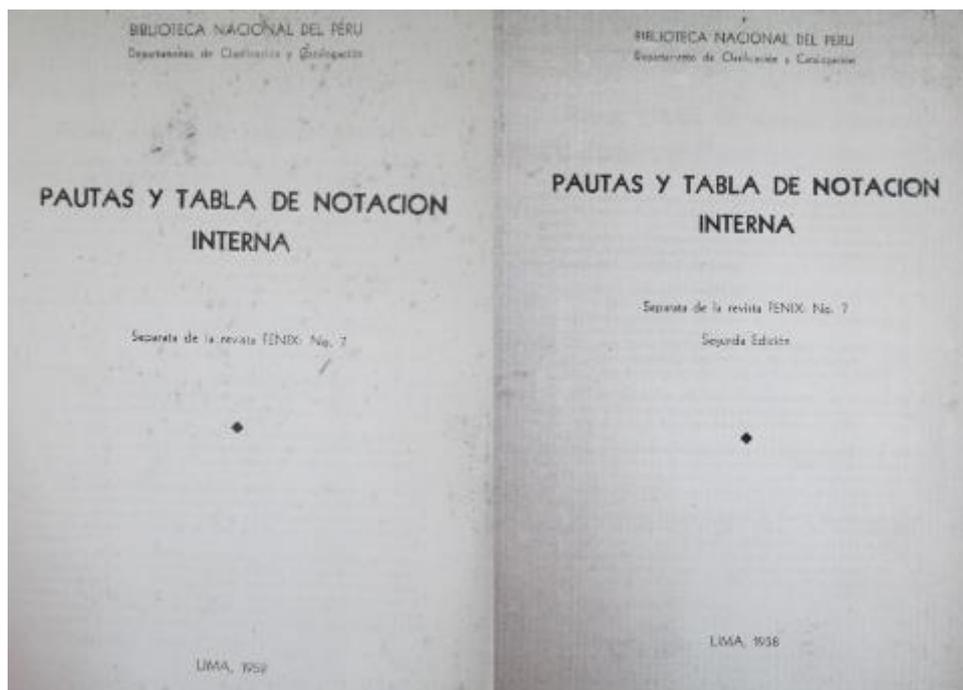


Figura 3. Vista de la página de inicio de la tabla de Málaga.
Fuente: Mundobiblio

2.2. Marco teórico

2.2.1. Investigaciones relacionadas

Autor	Año	Título de Estudio	Objetivos	Método	Resultados
Palomino Yesibel & Andrango Ronnal	2015	“Implementación de un sistema de información ejecutiva utilizando inteligencia de negocios para la eficaz interpretación de indicadores de atención y afiliación en el Seguro Integral de Salud para la administración de la Red de Salud de Huarochiri.”.	“Implementar un sistema de información ejecutiva utilizando una tecnología de inteligencia de negocios para la eficaz interpretación de indicadores de atención y afiliación en el Seguro Integral de Salud para la administración de una red de salud.”	Metodología de Ralph Kimball y Metodología Hefesto.	“Con respecto al objetivo general de la investigación podemos concluir que la implementación del sistema de inteligencia de negocios ha proporcionado una eficaz interpretación de indicadores en base a las encuestas realizadas y los resultados obtenidos con referencia a la satisfacción de un 90% por parte de la administración en el uso del sistema y validación de la información.”.
Huichi Quequejan a William Oscar	2015	“Implementación de una plataforma de Business Intelligence para la toma de decisiones en un centro de salud”.	“Implementar una plataforma de Business Intelligence para la toma de decisiones en un centro de salud”	Metodología S.A.F. E.	La implementación de una plataforma de Business Intelligence basada en un conjunto de dashboards e indicadores de atención ambulatoria para la mejor toma de decisiones, se ha convertido en una de las necesidades primordiales de cualquier alto directivo o jefe de áreas de un establecimiento de salud. Así como el proceso de tomas de decisiones gana eficiencia en la forma en que se incorporen herramientas de análisis de información que ayuden en la identificación de problemas para luego dar una solución o a realizar posibles tendencias basadas en los resultados obtenidos por medio de una plataforma de BI.

Lima Conejo Johana Marisol	2015	“Inteligencia de negocios para la toma de decisiones del departamento de cartera de la Cooperativa Finander”.	Implementar la inteligencia de negocios que mejoró la toma de decisiones dentro del departamento de cartera de la Cooperativa Finander.	Metodología de desarrollo de software XP.	Se consiguió cumplir con el Objetivo General al Implementar la inteligencia de negocios que mejoró la toma de decisiones dentro del departamento de cartera de la Cooperativa Finander y además se cumplió con los objetivos específicos propuestos mediante la fundamentación teórica y bibliográfica de los procesos de Inteligencia de Negocios para la toma de decisiones, y el diagnóstico de las necesidades de la empresa con respecto al departamento de cartera que se identificaron con la obtención de los resultados de las entrevistas, se consiguió desarrollar el modelo con los requerimientos expuestos por la institución para así validar la propuesta cumpliendo con las expectativas de los usuarios.
Zegarra Fuentes Gustavo Fernando	2015	“Solución de inteligencia de negocios orientada a mejorar la toma de decisiones en las operaciones mineras de extracción y metalurgia de Hochschild Mining”.	“Brindar soporte a la toma de decisiones empleando una solución de inteligencia de negocios de las operaciones de extracción y metalurgia diseñada e Implementada a la medida de Hochschild Mining.”	Metodología de Bill Inmon y Metodología de Ralph Kimball	“Se evidenció que, en el proceso de toma de decisiones a nivel de la gerencia de operaciones mineras, el tener una alta disponibilidad de la información, así como una síntesis fiable de la misma sobre los procesos de extracción y metalurgia resultó en una mejora en la toma de decisiones que propició la consecución de metas.”
Chávez Silvia & Contreras Carmen	2017	“Implementación de business intelligence, utilizando la metodología de Ralph Kimball, para el proceso de toma de decisiones del área de ventas. Empresa Yukids”.	Determinar en qué medida la implementación de Business Intelligence, aplicando a la metodología Ralph Kimball, mejora el proceso de toma de decisiones en el área de Ventas de la Empresa Yukids.	Metodología de Ralph Kimball.	La implementación de una solución de Business Intelligence, dio como resultado la reducción significativa en un 95% del tiempo promedio empleado en extraer información de ventas, pasando de horas de espera (3 horas 17 min promedio) a minutos (3.13 minutos en promedio).

Fuente: Elaboración propia

2.2.2. Business Intelligence

Hoy en día la información es abundante, las diversas fuentes de acceso a ellas, también lo son. Ello se ve reflejado en los diferentes sistemas de información utilizadas en las empresas que conforme pasa el tiempo nos proveen mejoras en sus funcionalidades, no obstante, usualmente se encuentra carencias en la integración de los sistemas puesto que los datos se encuentran aislados, y no tienen el análisis de los datos con mayor profundidad. De todo ello surge la necesidad de tener métodos eficientes para la extracción y transformación de los datos de una organización y convertirlos en información valiosa para éste, tales como: el método de la imputación (el cual permite que los datos tiendan al medio), el método de la eliminación de los datos nulos y los datos vacíos, el método de que los datos categóricos se conviertan en numéricos. El Business Intelligence atiende la necesidad expuesta anteriormente. La definición de BI ha evolucionado conforme pasa el tiempo, ya en el año 2011 se definía: “Se entiende por Business Intelligence al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización” (Curto, 2011, p. 18). Para el año 2018 una nueva definición expresa: “Business Intelligence (BI) es un término general que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento” (Gartner, 2018).

De las definiciones anteriores podemos resumir que el Business Intelligence es una agrupación herramientas y metodologías que transforman los datos en información provechosa que permiten el análisis de la organización basado en este conocimiento y en consecuencia se ejecuta una mejor toma de decisiones dentro de la misma.

Palomino & Andrango (2015) mencionan que la implementación de un sistema de inteligencia de negocios (Business Intelligence) en el Seguro Integral de salud para la administración de la Red de Salud de Huarochirí ha aportado la eficaz interpretación de indicadores con una efectividad del 90% en relación del desarrollo de los reportes, obteniendo información en tiempo real para la asignación de los recursos.

Huichi (2015) señala que el implementar una plataforma de Business Intelligence en un centro de salud se logró reducir el tiempo de la obtención, procesamiento y análisis de

información, también se pudo dar soporte a los altos directivos y jefes de áreas al momento de realizar la toma de decisiones sobre el control de la gestión de atenciones ambulatoria.

Según Lima (2015) la implementación de una solución de inteligencia de negocios en el departamento de cartera de la Cooperativa Finander consiguió integrar datos y mostrar resultados que apoyaron la toma de decisiones según los requerimientos de la institución.

Zegarra (2015) describe que la implementación de una solución de inteligencia de negocios en la empresa Hochschild Mining disminuyó sustancialmente el tiempo para la elaboración de reportes de estados de operaciones mineras así como se logró la alta disponibilidad de información que ayudó la toma de decisiones a nivel de la gerencia que propició la consecución de las metas de la misma.

De lo citado con anterioridad, se concluye que el aplicar Business Intelligence a una organización, ayuda en la disminución del tiempo de la obtención y análisis de los datos, apoyando la toma de decisiones de esta.

2.2.3. Toma de decisiones

A diario se toman decisiones, sean buenas o malas éstas tienen un impacto aun así fueren pequeñas. El tomar buenas decisiones es consecuencia a muchas experiencias aprendidas en el pasado, en función a ello la mente trabaja pensando en los diferentes resultados obtenidos de la decisión tomada. Las empresas también toman decisiones durante toda su vida institucional, una mala decisión puede acarrear consecuencias negativas dentro de éste, por tanto, es de vital importancia las decisiones tomadas dentro de ella.

2.2.3.1. Proceso de toma de decisiones

La toma de decisiones consta de ocho pasos según Robbins (2014), los cuales se describe a continuación:

1. Identificación del problema

Todas las decisiones tienen como causa un problema, la cual es una diferencia entre el estado existente y el esperado. La identificación del problema es algo subjetivo porque lo que para un gerente sea problema, tal vez para el otro no lo sea. Realizar este paso

de manera eficaz es de vital importancia, sin embargo no es una tarea sencilla (Robbins & Coulter, 2014).

2. *Determinación de los criterios de decisión*

Se debe establecer los criterios de selección que son primordiales para resolver el problema. Cuando se ve forzado a tomar una decisión, se elige como guía cualquiera de estos criterios aunque estos no fueren definidos en términos explícitos (Robbins & Coulter, 2014).

3. *Ponderación de los criterios*

Los criterios primordiales no tienen el mismo grado de importancia, de modo que la persona que va a tomar la decisión debe ponderar éstos, otorgándoles un peso correcto en la decisión. Un método de ponderación sencillo es darle al criterio más primordial el valor de 10 y según a ello asignar ponderación a los demás usando ese patrón (Robbins & Coulter, 2014).

4. *Desarrollo de las alternativas*

En este paso se demanda que el responsable de la toma de decisiones liste las alternativas posibles para solucionar el problema, limitándose a listar aquellas sin una previa evaluación (Robbins & Coulter, 2014).

5. *Análisis de las alternativas*

Se evalúan cada una de las alternativas según los criterios definidos en el paso número dos. En caso de que una alternativa obtenga una calificación muy alta en todos los criterios, no es necesario considerar las ponderaciones porque cada alternativa representará de por sí la mejor opción (Robbins & Coulter, 2014).

6. *Selección de una alternativa*

Se elige la mejor alternativa o aquella que haya obtenido el puntaje más alto en la ponderación del paso 5 (Robbins & Coulter, 2014).

7. *Implementación de una alternativa*

En esta etapa se pone en práctica la decisión. En primer lugar, se pone en conocimiento a quienes sean afectados por la decisión y obtener su compromiso respecto a ello de manera que serán más proclives a respaldarla. Luego se reevalúa el entorno para

localizar cualquier cambio que pueda ocurrir, más aun si ésta tiene efectos a largo plazo (Robbins & Coulter, 2014).

8. Evaluación de la eficacia

Se evalúa el resultado obtenido para saber si ello solucionó el problema. En caso de que la evaluación revelara que la dificultad persiste, el gerente debe buscar en que falló y responder estas preguntas:

- ¿Se definió correctamente el problema?
- ¿Se cometieron errores cuando se evaluaron las alternativas?
- ¿La alternativa elegida fue la correcta pero su implementación fue deficiente?

Las respuestas a estas interrogantes podrían conducir a nuevamente ejecutar algún paso anterior o forzarlo a reiniciar todo el proceso (Robbins & Coulter, 2014).

Se puede resumir los pasos mencionados con anterioridad con la siguiente figura:

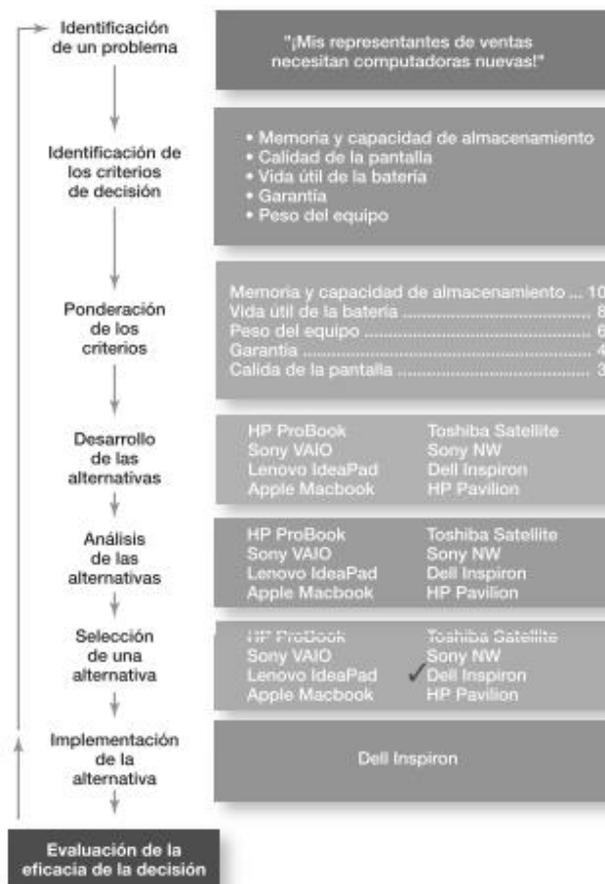


Figura 4. Proceso de toma de decisiones.
Fuente:(Robbins & Coulter, 2014).

2.2.3.2. Tipos de decisiones según su nivel

Se clasifican según el nivel administrativo del decisor o en función de la posición jerárquica de éste. Los cuales son los siguientes:

1. Estratégicas

La alta dirección de la organización se encarga de tomar las decisiones, éstas conciernen a los objetivos generales que afecta la organización en su totalidad, se caracterizan por ser de largo plazo y no se repiten. Las consecuencias que acarrearán estas decisiones son irreversibles, de manera que cometer un error compromete el progreso de la organización (Huichi, 2015).

2. Tácticas

Los directivos intermedios de la organización toman estas decisiones, se caracterizan por ser repetitivas, los errores no implican sanciones tan fuertes a no ser que éstas se vayan acumulando (Huichi, 2015).

3. Operativas

Los directivos de nivel más inferior de la organización toman estas decisiones, se caracterizan por estar relacionadas a las actividades cotidianas, su grado de repetición es muy alto a causa de que se basan en procedimientos automáticos. Los errores se pueden enmendar de manera rápida y las sanciones son mínimas (Huichi, 2015).

2.2.3.3. Tipos de decisiones según su método

1. Programadas

Se caracterizan por ser rutinarias cuando se tiene establecido un criterio para ello, ayudando a que no se trate de nuevo cada vez que se tome una decisión. No se decide en función a su grado de dificultad sino en función a su repetitividad así como su posibilidad de predicción y análisis de sus elementos (Huichi, 2015).

2. No programadas

Son nuevas para la empresa. No existe método definido para solucionar el problema ya que anteriormente no se había presentado, también porque su estructura es compleja o porque su relevancia es tan grande que merece una solución hecha a medida. Estas decisiones

se usan para problemas que puedan ocurrir de manera periódica pero que necesitan de una perspectiva modificada a causa de cambios de condiciones internas o externas (Huichi, 2015).

2.2.4. Metodología de Ralph Kimball

En la actualidad, las investigaciones respecto a la implementación de una solución de Business Intelligence, suelen utilizar como metodología para la construcción de un data warehouse o un datamart, la metodología de Ralph Kimball. A continuación, se mencionan las opiniones de los autores al usarla.

Palomino & Andrango (2015) señalan que al utilizar la metodología Ralph Kimball para el desarrollo de un datamart ayudó a determinar la mejora promedio que se aprecia es de un 50% en los procesos de interpretación de reportes de información en base a los indicadores de atención y afiliación, p. 90. Huichi (2015) menciona que la metodología más común y universal usada por las grandes empresas del desarrollo de plataforma de BI es Kimball, en donde muestra todo el proceso y entregables a realizar.

Birkbeck (2017) en su experiencia como consultor principal de BI para DATAMetrics Business Intelligence and Data Services, el cual está ubicado en Nueva Zelanda, describe que ha tendido a construir data warehouses casi exclusivamente con la Metodología de Ralph Kimball ya que es un marco para implementar una solución exitosa de Business Intelligence.

Por lo expuesto con anterioridad se concluye que la metodología de Ralph Kimball es la más usada en la construcción de data warehouses y datamarts, permitiendo que la solución de Business Intelligence sea eficiente y exitosa.

2.2.5. Matriz de similitud

Una matriz de similitud representa la distancia entre dos conjuntos de datos (Couprie, s. f.). Netflix es la compañía de alquiler con el mayor catálogo de películas en Estados Unidos, que usa la similitud de gustos de sus usuarios y según a ello recomienda las películas (Artés, 2011). Pandora es un sistema de recomendación de música, los usuarios que la usan proporcionan información a ésta acerca de sus canciones y artistas favoritos, ésta busca la similitud entre los usuarios y luego ésta les recomienda canciones cuyas

características almacenadas concuerden con sus preferencias (Artés, 2011). En las redes sociales se observa que también se usa la matriz de similaridad para mostrar resultados en las búsquedas hechas, según a los gustos e intereses del usuario.

CAPÍTULO III.METODOLOGÍA

3.1. Descripción del lugar de ejecución

Se ejecutó en el Centro de Recursos del Aprendizaje y la Investigación (CRAI) de la Universidad Peruana Unión.

3.2. Materiales e insumos

Tabla 7

Materiales e insumos

Herramientas	Tipo
Tableau	Licencia académica
Google Data Studio	Licencia académica
Python	Licencia PSFL (Python Software Foundation License, licencia de Software permisiva)
Spyder	Licencia MIT (Licencia de Software permisiva)
Kettle	Apache License

Fuente: Elaboración propia.

3.3. Metodología

3.3.1. Tipo de investigación

Diagnostica: Porque al establecer las características de los alumnos, el jefe del CRAI determinará las acciones que se deben tomar en el área para que ésta sea más eficiente con los alumnos que la visitan.

Predictiva: Porque la investigación tiene como propósito anticiparse ante situaciones futuras.

Explicativa: Porque la investigación estudia el porqué del hecho.

Propositiva: Porque la investigación propone una solución al problema de obtener reportes interactivos.

Aplicativa: Porque la investigación da respuesta a objetivos específicos.

3.3.2. Arquitectura de solución

Escobar & Gaete (2015) en su investigación “Desarrollo de una solución de inteligencia de negocios, para procesos mineros” señala como arquitectura de solución una capa de datos en la cual está almacenada la data de la organización, una capa de extracción en la cual se normaliza los datos descritos, una capa de QVD y una capa de presentación de los datos (p. 33). A continuación, se muestra esta arquitectura en la *Figura 5*.

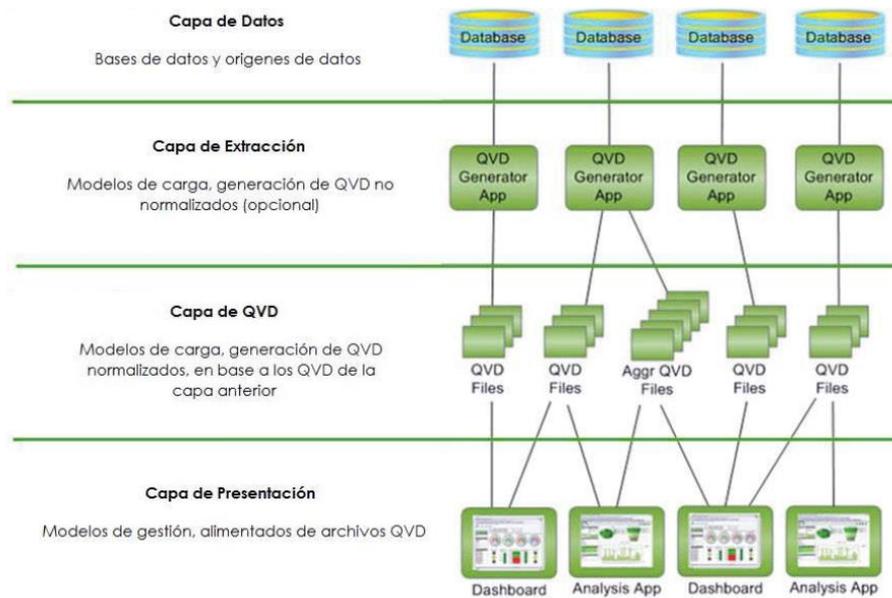


Figura 5. Arquitectura de solución BI de Escobar & Gaete.
Fuente: (Escobar & Gaete, 2015, fig. 4.7).

Contel (2011) en su investigación “Desarrollo de una solución Business Intelligence en una empresa del sector de alimentación” indica como arquitectura de solución una data histórica del Sistema Operacional SAGE que pasa por un proceso de Extracción, Transformación y Carga (ETL) para cargarlo a un datawarehouse y subirlo a la nube y poder mostrar los dashboards y reportes, tal como lo muestra la siguiente figura:

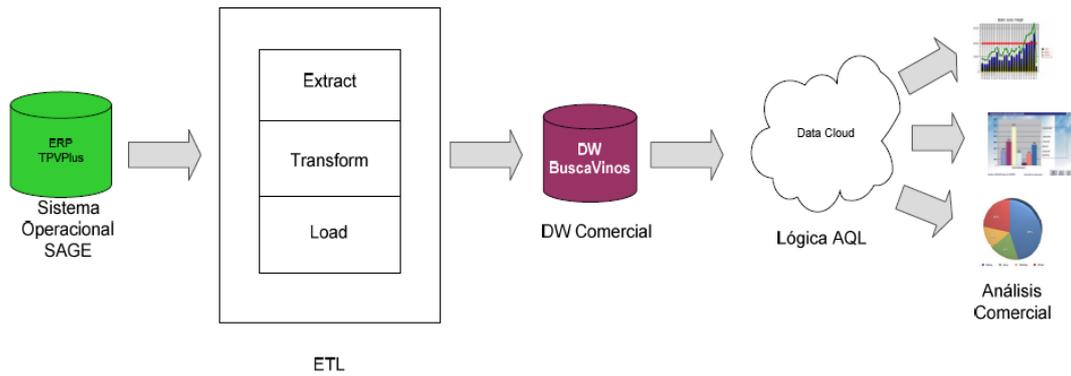


Figura 6. Arquitectura de solución de BI de Contel.

Fuente: (Contel, 2008, fig. 28).

Peña (2015) en su investigación “Diseño de una arquitectura de Inteligencia de Negocios para el área de compras de seguros Bolívar” muestra como arquitectura de solución BI un origen de datos que son almacenados en un datawarehouse, del cual dentro de éste hay un datamart donde se hace el análisis de procesamiento en línea (OLAP) para mostrar reportes y dashboards que generen conocimiento y permitan la toma de decisiones, tal como se muestra la Figura 7.

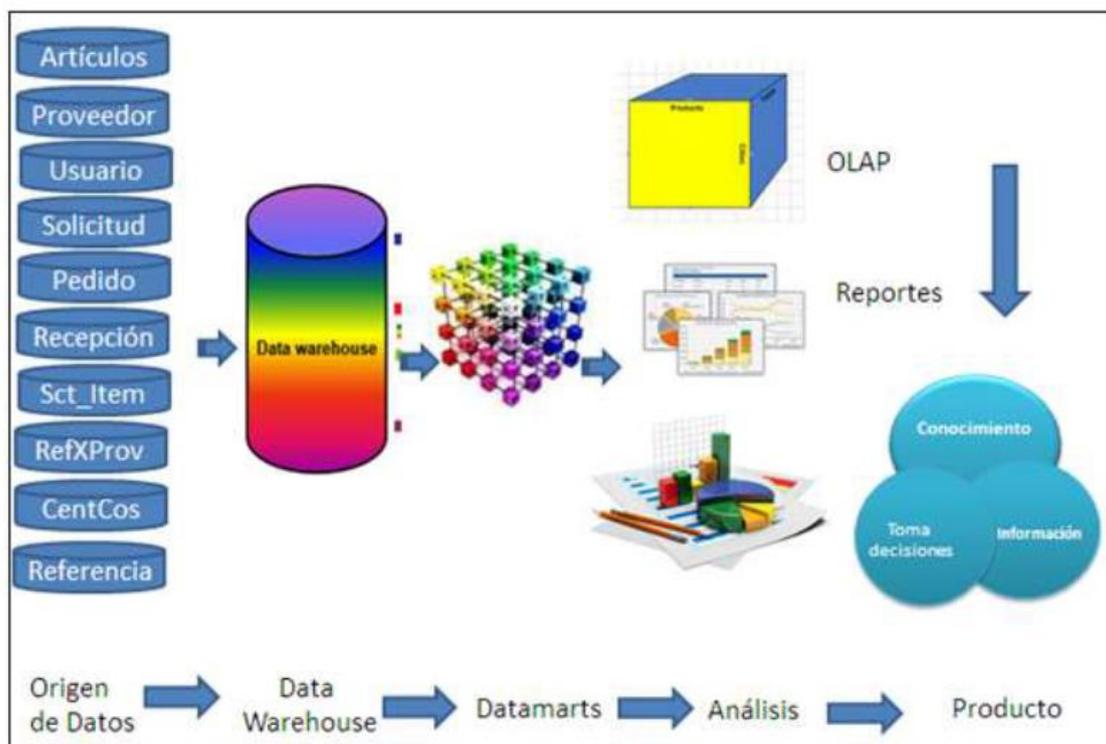


Figura 7. Arquitectura de solución BI de Peña.

Fuente: (Peña, 2015, fig. 26).

De las investigaciones descritas con anterioridad, tomando en cuenta los orígenes de datos, el proceso ETL, la carga de datos y la obtención de dashboards es que el investigador ha podido crear su propia arquitectura de solución BI, basada en ellas. La arquitectura de solución se ve reflejada en la siguiente figura:

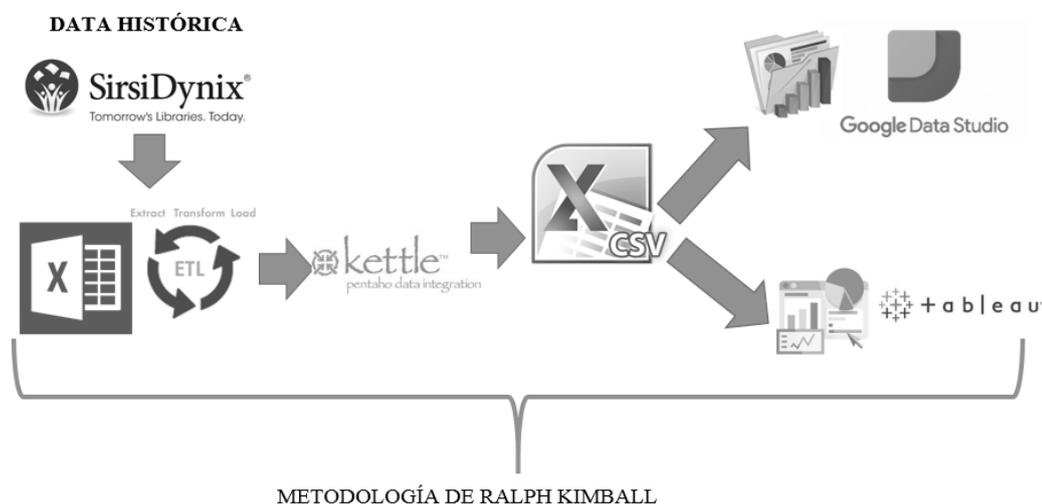


Figura 8. Arquitectura de solución de la plataforma BI.

Fuente: Elaboración propia basada en la metodología de Ralph Kimball.

Como podemos observar en la *Figura 8* se extrajeron los datos históricos del sistema SirsiDynix, los cuales están en archivos Excel, con formato xlsx. Luego de ello se procedió con la limpieza de los datos, lo cual fue de manera manual, usando funciones del mismo Excel para este proceso. Posteriormente se transformó los datos hacia el formato csv con la herramienta Kettle; y por último se realizó los reportes con dos herramientas: Google Data Studio y Tableau para hacer una comparación entre éstas. Todos los pasos que se sigan en la arquitectura de solución se harán conforme a la metodología de Ralph Kimball.

Se ha visto conveniente usar la herramienta Kettle para la transformación y limpieza de datos sobre TalendETL a causa de que Kettle es muy intuitiva, de mejor performance y manejo de errores.

TalendETL respecto a su performance es más lento ya que se demora en la carga de código Java generado. En relación al manejo de errores, si el trabajo falla debe solucionarse el problema y reiniciarse el proyecto; es poco intuitiva y difícil de entender. Es un generador de código y ello hace que se dependa del lenguaje elegido en los proyectos (Java o Groovy) necesitando un nivel alto de éste para tener la mayor potencia de ésta herramienta.

Se usó Kettle porque la herramienta es muy intuitiva y con unos conceptos básicos se puede empezar a ejecutar un proyecto. Su motor de transformación, ha sido diseñado con la finalidad de cubrir la necesidad de la integración de datos. Respecto a su performance es más rápida ya que no tiene que estar moviendo código generado.

Se utilizó Tableau y Google Data Studio sobre Qlikview (La interfaz y la facilidad de uso requieren una curva de aprendizaje un poco larga para aprender sus funcionalidades. Su secuencia de comandos funciona como ETL. Costo elevado por sus licencias.), y MicroStrategy (precio elevado ya que está enfocado más a su implementación en grandes empresas), siendo que por sobre las demás estas presentan una interfaz intuitiva y amigable, tienen la capacidad de ejecutar cálculos complejos, así como el análisis integrado avanzado de los datos. Otra de las razones es que las herramientas seleccionadas nos permiten acceder a ella de manera libre de pago, por ejemplo, se puede obtener Google Data Studio con solo una cuenta de correo de Google y respecto a Tableau nos otorga una licencia a los estudiantes universitarios para acceder a su herramienta. Asimismo, se eligió Tableau porque lidera el Cuadrante de Gartner, el cual es un gráfico donde se ubican las principales empresas proveedoras de servicios tecnológicos del mundo, el objetivo principal del cuadro es mostrar los productos impulsan la mayoría de las compras en el mercado actual que cumplen con los criterios de una plataforma de analytics y BI.

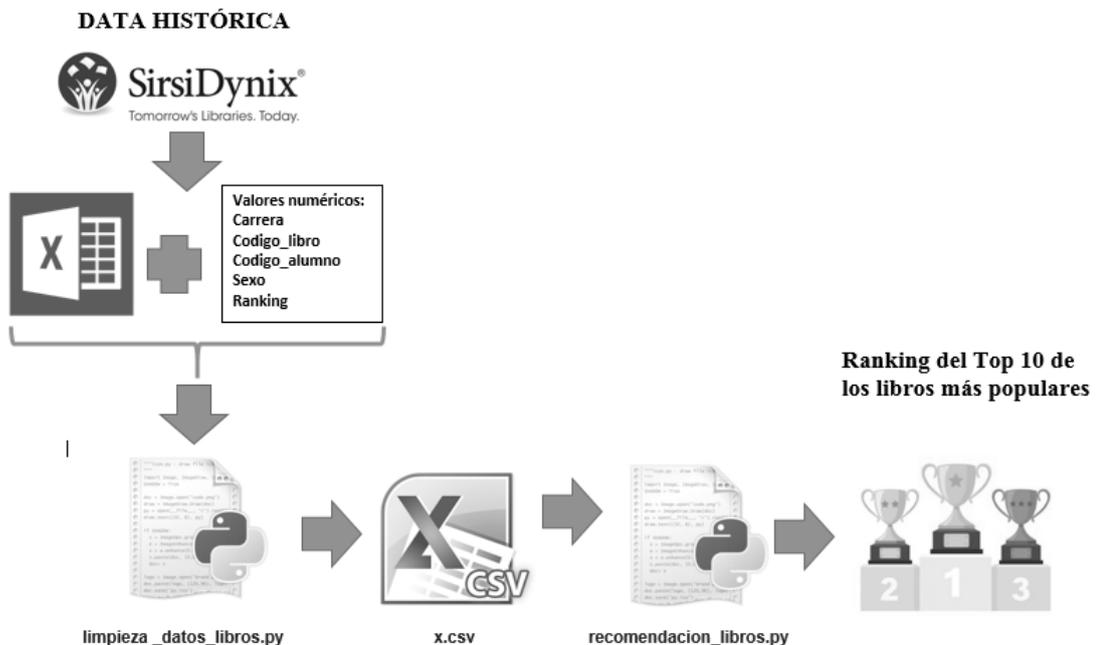


Figura 9. Arquitectura de solución del sistema de recomendación.
Fuente: Elaboración propia.

La *Figura 9* muestra la arquitectura de solución del sistema de recomendación de libros, en el cual se numerará las carreras con un código, lo mismo con los libros, el alumno, el sexo; también se le puso un ranking a cada alumno, siendo el valor 1 malo, el valor 2 regular, el valor 3 bueno y el valor 4 excelente. Se programó un código de limpieza en Spyder con los paquetes que ofrece PyData, que permite la limpieza de las columnas que no utilizamos, así como el reemplazo de los valores nulos o vacíos con la media. Luego de ello como resultado se obtendrá un archivo con extensión csv, con las columnas necesarias para cargar a otro código programado que muestra el top 10 de los libros más usados.

Se usó Spyder por sobre Jupyter (proceso complejo de instalación) porque IDE simple y liviano de fácil instalación, con documentación detallada. Es un editor de código abierto. Compatible con la depuración en tiempo de ejecución, esto quiere decir que los errores se ven en pantalla tan pronto como se los escribe. Se integra con alguna de las librerías fundamentales, tales como, NumPy, Matplotlib, SciPy, entre otras.

También se utilizó PyData porque una de las suites gratuitas que nos ofrece es ANACONDA, la cual es una distribución de Python que contempla diferentes aplicaciones y librerías para el desarrollo de ciencia de datos con Python, ésta contiene más de 1000 paquetes de datos.

3.4. Metodología de Ralph Kimball

Para desarrollar la presente investigación se empleó la metodología de Ralph Kimball, la cual se describió en el marco conceptual.

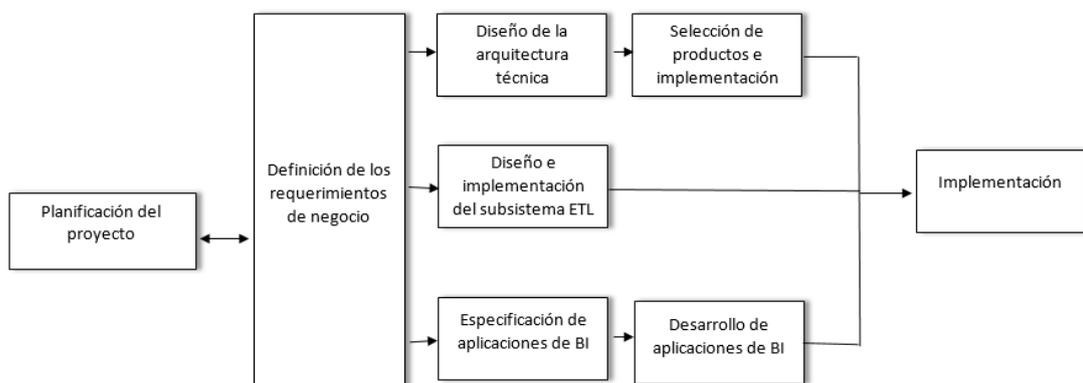


Figura 10. Metodología de Ralph Kimball.
Fuente: Elaboración propia.

A continuación, se presentan las actividades realizadas en cada etapa de la metodología:

3.4.1.1. Planificación del proyecto

Diagnóstico de la empresa: Es la primera etapa del proyecto, en la cual se realizó el levantamiento de información del CRAI mediante entrevistas.

Plan de proyecto: En esta etapa se definió las actividades, el alcance y el cronograma del proyecto incluyendo recursos y tiempo.

- Alcance: La implementación de la plataforma BI en el CRAI sólo abarca el proceso de préstamo de literatura usando la metodología de Ralph Kimball para la toma de decisiones del área.
 - Roles del proyecto: Se concedieron roles a los participantes involucrados de manera que se tenga en claro cuál es su función de cada uno de ellos; como patrocinador del proyecto estuvo el CRAI a través de su Director, como Director del proyecto estuvo la Bach. Narda Naomi Lanuza Bustamante.
- Actividades y cronograma del proyecto: Se establecieron las actividades con su respectiva duración de manera que el proyecto esté organizado, tal como se muestra en la *Tabla 8*.

Tabla 8

Cronograma del proyecto

Nombre de tarea	Duración	Comienzo	Fin
Proyecto de Implementación de BI	64 días	jue 01/11/18	mar 29/01/19
Planificación del proyecto	17 días	jue 01/11/18	vie 23/11/18
Diagnóstico del CRAI	8 días	jue 01/11/18	lun 12/11/18
Realizar el plan de proyecto	9 días	mar 13/11/18	vie 23/11/18
Análisis de requerimientos del negocio	13 días	sáb 24/11/18	mié 12/12/18
Definir los requerimientos	7 días	sáb 24/11/18	lun 03/12/18
Definir la arquitectura	7 días	mar 04/12/18	mié 12/12/18
Diseño de la arquitectura técnica	13 días	jue 13/12/18	lun 31/12/18
Selección de productos e implementación	7 días	jue 13/12/18	vie 21/12/18
Diseño e implementación del subsistema ETL	7 días	sáb 22/12/18	lun 31/12/18
Especificación de aplicaciones de BI	7 días	mar 01/01/19	mié 09/01/19
Desarrollo de aplicaciones de BI	7 días	jue 10/01/19	vie 18/01/19
Implementación	7 días	lun 21/01/19	mar 29/01/19

Fuente: Elaboración propia.

- Estructura de Desglose de Trabajo: Para una adecuada organización del proyecto, se construyó la EDT a partir del cronograma establecido, tal como lo presenta la *Figura 11*.

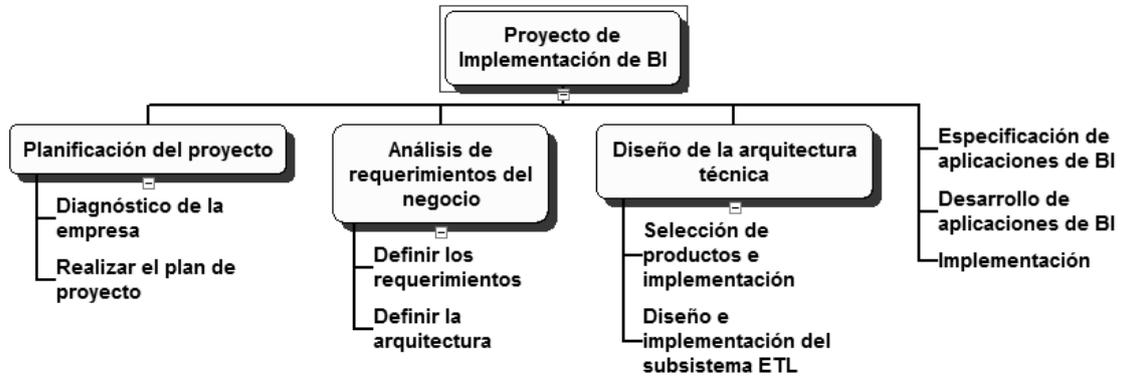


Figura 11. Estructura de Desglose de Trabajo del proyecto.

Fuente: Elaboración propia.

3.4.1.2. Análisis de requerimientos de negocio

Definición de Requerimientos: Se realizó entrevistas al jefe del CRAI: Ing. Rogelio Mamani, al jefe de Tecnologías de la Información del área en mención: Ing. Walter Luque; lo cual permitió entender los procesos del negocio, tales como: proceso de préstamo, devolución y adquisición de recursos de estudio (libros). De los cuales en la *Figura 12* pongo como evidencia uno de los procesos que fue seleccionado para implementar la plataforma de inteligencia de negocios.

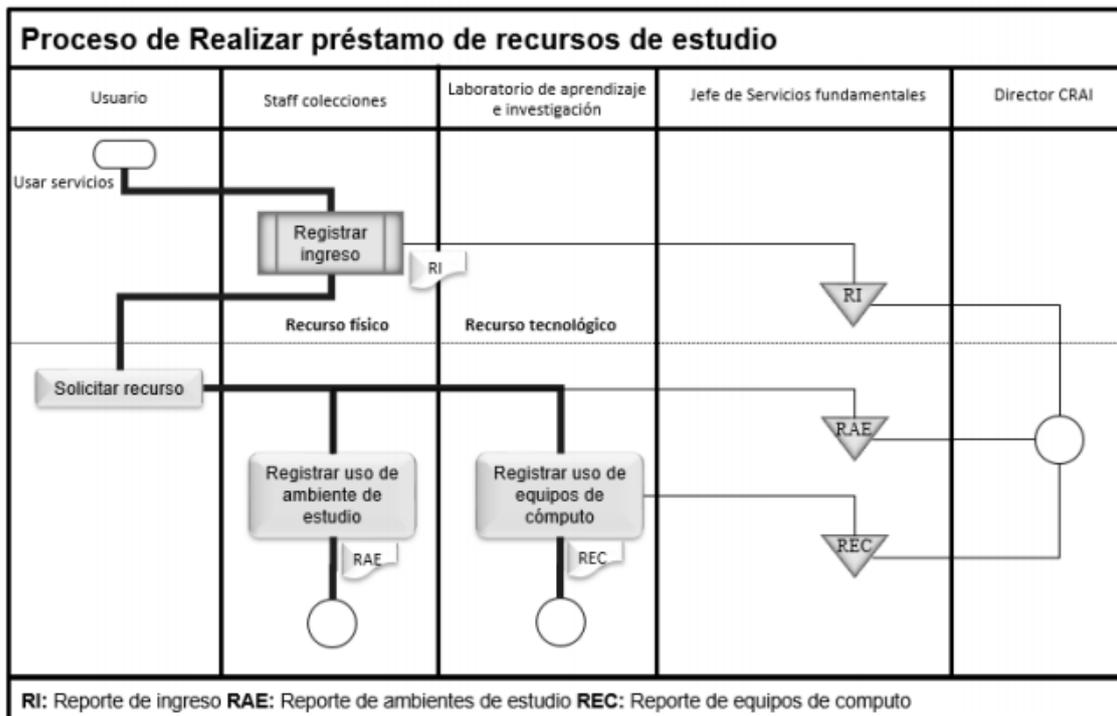


Figura 12. Proceso de préstamo de libro.
Fuente: Procesos UPeU.

Con ello también se logró comprender los probables requerimientos; en la *Tabla 9* se muestra la lista de requerimientos funcionales y en la *Tabla 10* se muestra los requerimientos no funcionales.

Tabla 9

Requerimientos funcionales

ID	REQUERIMIENTO	DESCRIPCIÓN
1	Visualizar las carreras que se prestan más libros	El informe debe mostrar un indicador que presente el ranking de las carreras profesionales según la cantidad de préstamo de libros. Se debe permitir hacer combinación y agrupación de datos sobre el conteo de préstamos y la unidad de tiempo a fin de obtener información más estratégica.
2	Visualizar la duración del préstamo de libros según sus diferentes ítems.	Se debe mostrar la duración del préstamo según a los ítems: nombre del libro, datos del alumno, carrera o escuela profesional u otro que se estime por conveniente.
3.	Visualizar la cantidad de préstamos de libros según los años.	Se debe permitir el filtro de datos de acuerdo a los años y carrera asociada al estudiante.
4	Tener opción de visualizar pronósticos y tendencias.	Se debe permitir la visualización de pronósticos de manera que ayude a la toma de decisiones estratégicas en el área.

Fuente: Elaboración propia.

Tabla 10

Requerimientos No Funcionales

ID	TIPO	DESCRIPCIÓN
1	Disponibilidad	Es necesario tener disponibilidad inmediata de la información, se debe contar siempre con la información relacionada a la última transacción finalizada a nivel de base de datos.
2	Desempeño	Se espera que el análisis de la información sea lo más eficaz posible.
3	Simplicidad	El sistema debe ser fácil de usar
4	Interfaz del software	Compatibilidad con archivos Excel.

Fuente: Elaboración propia

Definición de la arquitectura: La arquitectura en donde se enmarcó la configuración de toda la solución, se planteó y justificó según las prioridades evaluadas en los requerimientos funcionales y no funcionales planteados.

3.4.1.3. Diseño de la arquitectura técnica

Se identificó las herramientas para satisfacer los requerimientos de negocios, tal como lo muestra la *Figura 8* de la arquitectura de solución.

3.4.1.4. Selección de productos e implementación

Se propuso un conjunto de herramientas como: Kettle, Tableau, Google Data Studio; que pudo satisfacer las necesidades específicas cuyo costo de la licencia fue cero. La elección de herramientas comprendió varios niveles para cada uno de los niveles de implementación.

A continuación, se detallan en la siguiente tabla:

Tabla 11

Softwares de instalación

Producto	Característica	Uso
Pentaho Data Integration (Kettle)	Integrador de datos en la Extracción, Transformación y carga de datos (ETL)	Convertir el archivo xlsx a formato csv
Tableau	Plataforma de Business Intelligence	Diseño y construcción de dashboards, reportes.
Google Data Studio	Plataforma de Business Intelligence	Diseño y construcción de dashboards, reportes.

Fuente: Elaboración propia.

3.4.1.5. Diseño e implementación del subsistema ETL

Se extrajeron los datos del software usado en el CRAI el cual es: SirsiDyNix Symphony tal como lo muestra la *Figura 13*, en un formato xlsx.

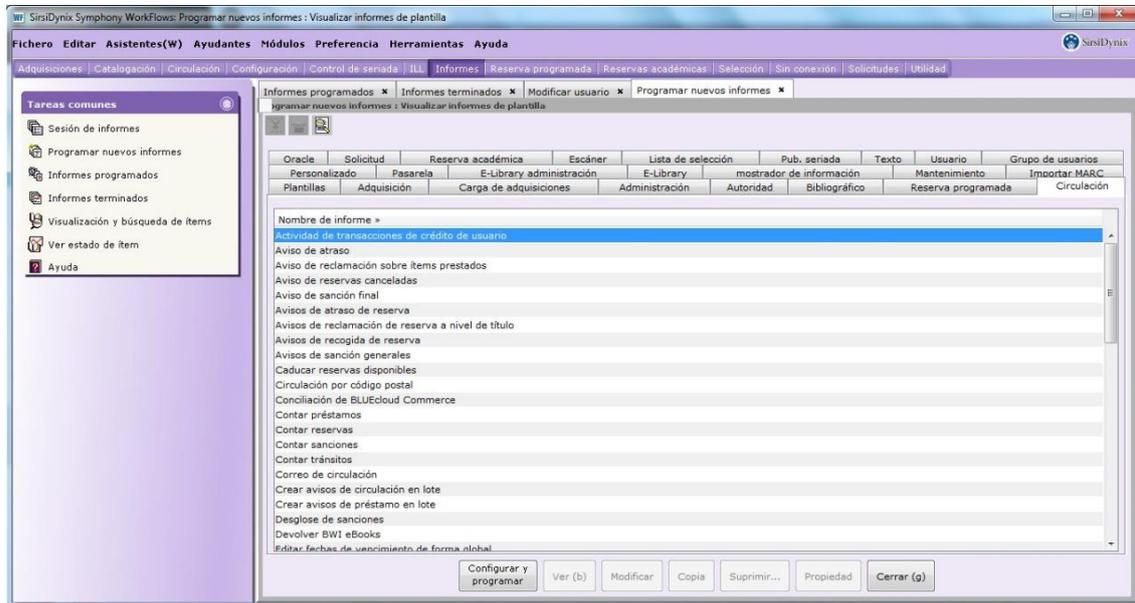


Figura 13. Vista del panel de exportación de informes SirsiDyNix Symphony
Fuente: SirsiDyNix CRAI

La data en bruto, sin limpieza de datos se muestra a continuación:

	A	B	C	D	E	F	G
	NombreS	Apellidos	Libros prestados	Fecha de préstamo	Fecha de de	Sexo	Carrera
1	Braulio	Gutierrez Pari					
2							CIVIL
3						VARON	
4				28/08/18	04/09/18		
5	Walter	Luque Condori	La ingeniería de suelos en las vías terrestres : carreteras, ferrocarriles y				
6							SISTEMAS
7						VARON	
8				21/08/18	23/08/18		
9	Gabriel	Vela Vásquez	Contabilidad general : Teoría y práctica / Lázaro Rodríguez Ariza, María V				
10							TEOLOGIA
11						VARON	
12				23/08/18	31/08/18		
13			El poder de la bendición paternal / Graig Hill.	23/08/18	31/08/18		
14			Momentum : el poder de la herencia generacional. Lo que Dios comien:	23/08/18	31/08/18		
15			La nueva vida después del divorcio / Bill Butterwort.	23/08/18	31/08/18		
16			Viviendo la tercera edad : un modelo integral de consejería para el buei	23/08/18	31/08/18		
17	Segundo	Azo Salazar					
18							TEOLOGIA
19						VARON	
20			Mateo : la revelación de la realeza de Cristo, Mateo 15-28 / Evis L. Carbal	26/08/18	26/08/18		
21			Mateo : la revelación de la realeza de Cristo, Mateo 1-14 / Evis L. Carball	26/08/18	02/09/18		
22	Isaac	Sáenz Mori					
23							ARQUI

Figura 14. Data histórica en bruto.
Fuente: Elaboración propia.

Los cuales se limpiaron con las funciones de Excel tales como: CONTAR.SI.CONJUNTO, el cual se muestra en la *Figura 15*; función TRUNCAR que se

muestra en la *Figura 16*; función SI que se muestra en la *Figura 17*; función SUMAR.SI.CONJUNTO que se muestra en la *Figura 18*.

M2 : *fx* =CONTAR.SI.CONJUNTO(\$C\$2:\$C\$466,D2)

	C	D	F	G	H	I	J	K	L	M
1	Apellidos	Apell	Lit	Fecha de pr	Fecha de	Sexo	Carrera	DURACIÓN	RANKING	VECES PRESTADO POR ALUMNO
2	Gutierrez Pari	Gutierr	La ing	28/08/2018	04/09/2018	VARON	CIVIL	7	EXCELENTE	1
3	Luque Condori	Luque C	Contá	21/08/2018	23/08/2018	VARON	SISTEMAS	2	BUENO	1
4	Vela Vásquez	Vela Vá	El por	23/08/2018	31/08/2018	VARON	TEOLOGIA	8	EXCELENTE	4

Figura 15. Función Excel CONTAR.SI.CONJUNTO.
Fuente: Elaboración propia.

K2 : *fx* =TRUNCAR(H2-G2,0)

	C	D	F	G	H	I	J	K
1	Apellidos	Apell	Lit	Fecha de pr	Fecha de	Sexo	Carrera	DURACIÓN
2	Gutierrez Pari	Gutierr	La ing	28/08/2018	04/09/2018	VARON	CIVIL	7

Figura 16. Función Excel TRUNCAR.
Fuente: Elaboración propia.

L2 : *fx* =SI(K2<5,"BUENO","EXCELENTE")

	C	D	F	G	H	I	J	K	L
1	Apellidos	Apell	Lit	Fecha de pr	Fecha de	Sexo	Carrera	DURACIÓN	RANKING
2	Gutierrez Pari	Gutierr	La ing	28/08/2018	04/09/2018	VARON	CIVIL	7	EXCELENTE

Figura 17. Función Excel SI.
Fuente: Elaboración propia.

1	Apellidos	Apel	Li	Fecha de pr	Fecha de i	Sexo	Carrera	DURACIÓN	RANKING	VECES PRESTADO POR ALUMNO
2	Gutierrez Pari	Gutierr	La ing	28/08/2018	04/09/2018	VARON	CIVIL	7	EXCELENTE	1

Figura 18. Función Excel SUMAR.SI.CONJUNTO.

Fuente: Elaboración propia.

Para añadir más datos que nos ayuden a comprender las características de cada usuario presentado en la base de datos, se buscó en el Sistema Académico los siguientes datos: la sede a la cual pertenece y el tipo de usuario: alumno o docente; a continuación se muestra en la *Figura 19* el menú principal del sistema académico, en la *Figura 20* se muestra la búsqueda de los datos del alumno y en caso de que el usuario no se encontrara en la búsqueda del alumno se hizo la búsqueda los datos de docente que se muestra en la *Figura 21*.

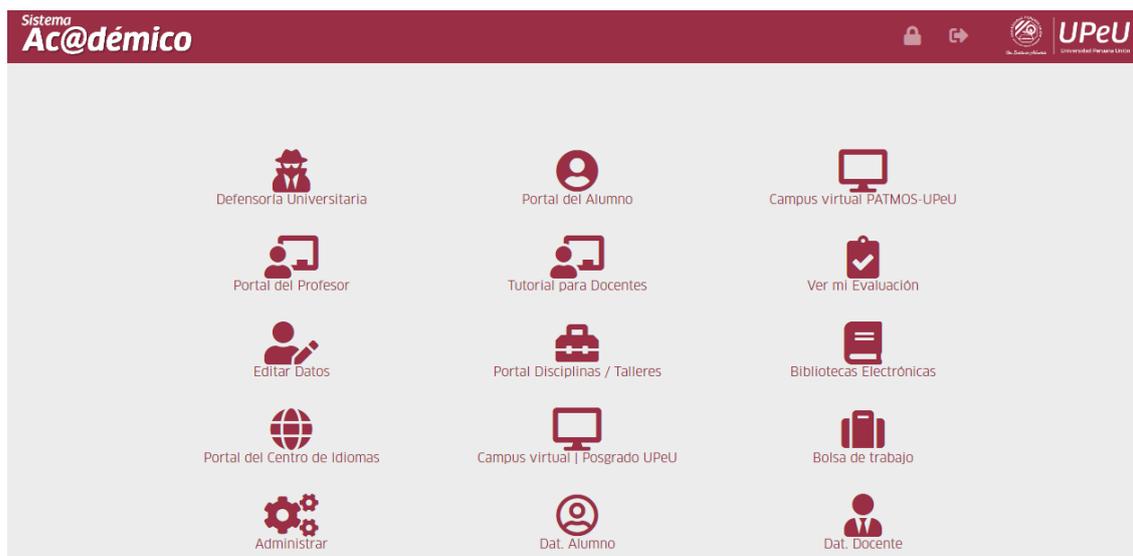


Figura 19. Menú principal del sistema académico.

Fuente: Elaboración propia.

Figura 20. Vista de la búsqueda de los datos del alumno.
Fuente: Elaboración propia.

Figura 21. Vista de la búsqueda de datos del docente.
Fuente: Elaboración propia.

Al finalizar la limpieza nuestra data quedó de la siguiente manera, como lo muestra la figura:

	B	C	D	E	F	G	H	I	J	K	L
1	Apellidos	Nombre S	Carrera	Fecha de devolución	Fecha de préstamo	Codigo de Lit	Nombre_libro	Ranking	Sexo	SEDE	TIPO
2	Gutiérrez Pari	Braulio	CIVIL	04/09/2018 21:30	28/08/2018 14:34	1	La ingeniería de suelos en las vías ter	EXCELENTE	VARON	JULIACA	DOCENTE
3	Luque Condori	Walter	SISTEMAS	29/08/2018 21:30	21/08/2018 09:13	2	Contabilidad general : Teoría y prácti	BUENO	VARON	JULIACA	DOCENTE
4	Vela Vásquez	Gabriel	TEOLOGIA	31/08/2018 21:30	23/08/2018 16:21	3	El poder de la bendición paternal / Gi	EXCELENTE	VARON	LIMA	DOCENTE
5	Vela Vásquez	Gabriel	TEOLOGIA	31/08/2018 21:30	23/08/2018 16:21	4	Momentum : el poder de la herencia	EXCELENTE	VARON	LIMA	DOCENTE
6	Vela Vásquez	Gabriel	TEOLOGIA	31/08/2018 21:30	23/08/2018 16:21	5	La nueva vida después del divorcio /	EXCELENTE	VARON	LIMA	DOCENTE
7	Vela Vásquez	Gabriel	TEOLOGIA	31/08/2018 21:30	23/08/2018 16:21	6	Viviendo la tercera edad : un modelo	EXCELENTE	VARON	LIMA	DOCENTE
8	Azo Salazar	Segundo	TEOLOGIA	26/08/2018 20:24	26/08/2018 18:24	7	Mateo : la revelación de la realeza de	MALO	VARON	LIMA	DOCENTE
9	Azo Salazar	Segundo	TEOLOGIA	02/09/2018 21:30	26/08/2018 18:30	8	Mateo : la revelación de la realeza de	EXCELENTE	VARON	LIMA	DOCENTE
10	Sáenz Mori	Isaac	ARQUI	09/12/2015 21:30	26/11/2015 20:28	9	Introducción a la literatura esenia de	EXCELENTE	VARON	LIMA	DOCENTE
11	Aguirre Zambrano	Romeen	POSGRADO	31/08/2018 21:30	27/08/2018 18:42	10	Curso práctico de técnicas de comuni	BUENO	VARON	LIMA	DOCENTE
12	Rojas Atanacio	Leonardo	TEOLOGIA	16/01/2018 21:30	09/01/2018 14:36	11	Lecciones de economía / Carlos Bolo	EXCELENTE	VARON	LIMA	DOCENTE
13	Choque Fernández	Juan	POSGRADO	17/01/2018 21:30	10/01/2018 11:07	12	Iniciación Filosófica / Augusto Salaz	EXCELENTE	VARON	LIMA	DOCENTE
14	Arana Rodríguez	Maritza	ADMIN	17/01/2018 21:30	10/01/2018 11:08	13	Breve Antología Filosófica / Augusto	EXCELENTE	MUIJER	LIMA	DOCENTE
15	Meza Silva	Alejandro	TEOLOGIA	31/08/2018 21:30	23/08/2018 14:09	14	Administración estratégica / Alfredo	EXCELENTE	VARON	LIMA	ALUMNO

Figura 22. Base de datos limpia.
Fuente: Elaboración propia.

La transformación se realizó con Kettle (Pentaho Data Integration), en este paso se transformó del formatoxlsx al csv, tal como lo muestra la Figura 23.

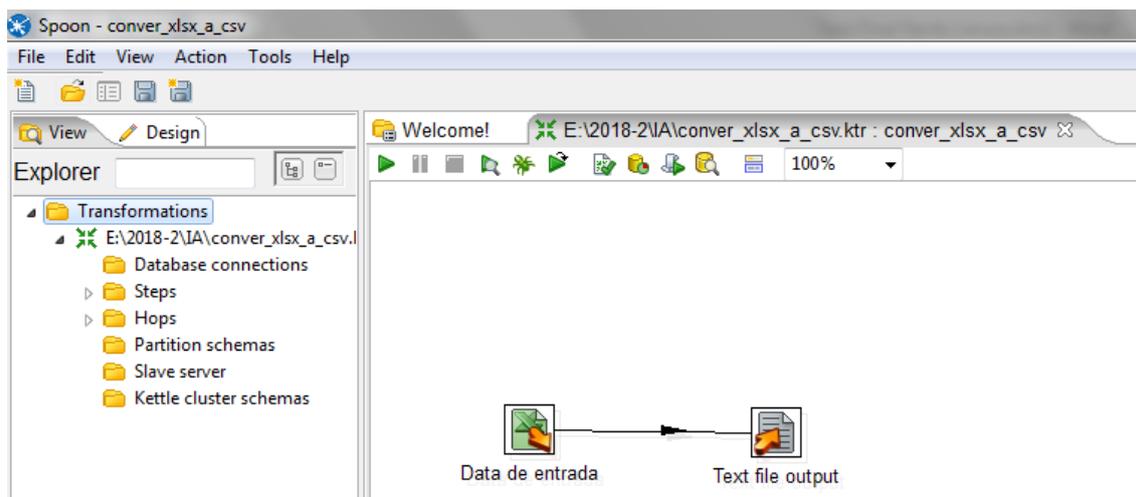


Figura 23. Conversión de archivo a xlsx a csv en Pentaho Data Integration (Kettle).
Fuente: Elaboración propia.

Los datos finales estuvieron listos para cargar al Tableau Professional Edition y al Google Data Studio.

3.4.1.6. Especificación de aplicaciones de BI

Se definió las aplicaciones analíticas de BI a usarse y los reportes de una manera detallada según el rol o el perfil del usuario, en este caso el nivel de análisis de los reportes, fueron para los jefes del CRAI. De los cuales en la *Tabla 12* se muestra a detalle el contenido de ello.

Tabla 12

Reportes de visualización de datos

ID	REPORTE	DESCRIPCIÓN	TIPO DE DECISIÓN EN LA QUE AYUDA EL REPORTE
1	Tendencia del comportamiento de los préstamos de las carreras según el mes del año	Muestra los meses del año y la cantidad de préstamos hechos en cada uno de ellos por cada carrera.	Estratégica
2	Cantidad de préstamos total según cada mes del año	Muestra los meses del año y la cantidad de préstamos hechos de manera global.	Táctica
3.	Pronóstico de la cantidad de préstamos total según cada mes del año	Muestra el pronóstico de préstamos según el mes del año.	Estratégica
4	Promedio de la cantidad de préstamos hechos según la hora del día según el año	Muestra la cantidad de préstamos según la hora del día y el año.	Táctica

5	Cantidad de préstamo total de los estudiantes según su sexo	Muestra la cantidad en porcentaje del total préstamos según su sexo	Táctica
6	Duración de préstamos según la carrera	Muestra la duración de días de préstamos según las carreras	Táctica
7	Duración de préstamos de libros según carrera y sexo	Muestra la duración de días de préstamos según las carreras y sexo.	Táctica
8	Clustering (agrupamiento) de alumnos según la cantidad de días de préstamos de libros	Muestra la agrupación de alumnos según la cantidad de días de préstamo en 3 clústers.	Táctica
9	Top 10 de los libros más recomendados	Muestra el top 10 de los libros más recomendados mediante el algoritmo KNN,	Estratégica

Fuente: Elaboración propia

3.4.1.7. Desarrollo de aplicaciones de BI

1. Desarrollo del reporte de la línea de tendencia de comportamiento de los préstamos de las carreras según el mes del año

En primer lugar, al abrir el Tableau, nos muestra su panel de inicio, donde le damos clic donde dice conectar a un archivo Excel tal como lo muestra la siguiente figura:



Figura 24. Panel de inicio de Tableau.

Fuente: Elaboración propia.

Luego de ello damos clic donde está el archivo data_limpiar.csv y luego clic en ABRIR, como muestra la siguiente figura:

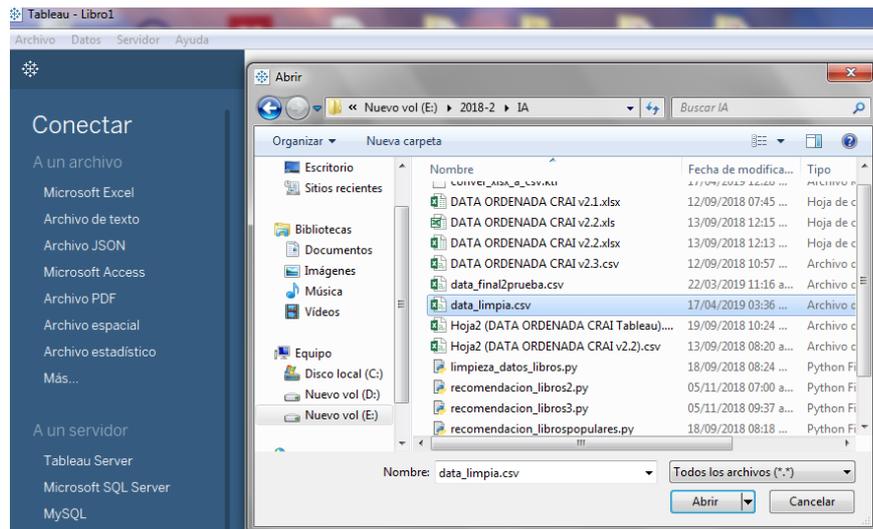


Figura 25. Carga del archivo limpio al Tableau.
Fuente: Elaboración propia.

Tableau redirigirá a un panel donde permite la creación de reportes; una parte del panel muestra las medidas y las dimensiones; las medidas son atributos cuantitativos, en este caso el código del libro, días de préstamo y número de registros; por otro lado las dimensiones son aquellos atributos cualitativos que describen las características de las medidas, en este caso son nombres (del alumno), nombre del libro, sede (Lima, Juliaca o Tarapoto), carrera, sexo (mujer o varón), tipo (alumno o docente), fecha de préstamo y fecha de devolución, como se muestra a continuación en la figura:

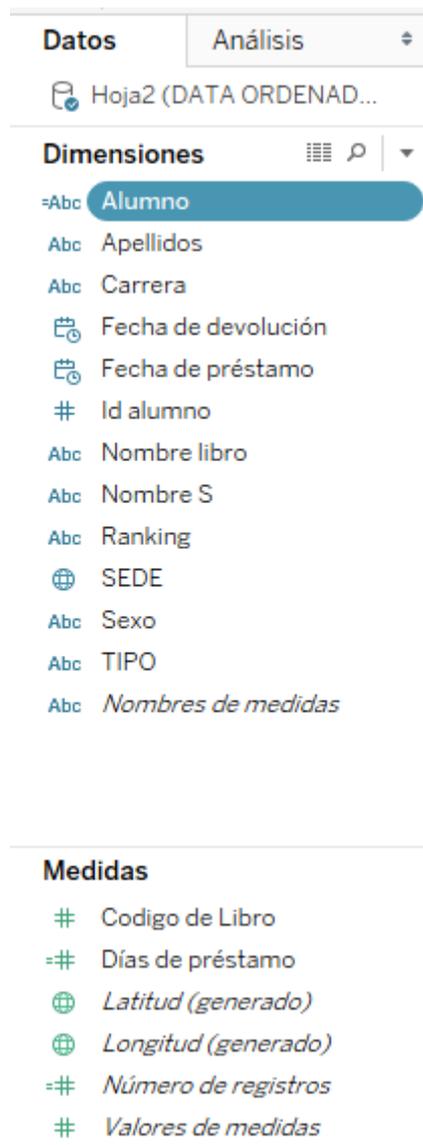


Figura 26. Panel de dimensiones y medidas del proyecto.
Fuente: Elaboración propia.

Para la configuración del reporte arrastramos la medida NUMERO DE REGISTROS hacia la barra de filas, arrastramos la dimensión FECHA DE PRÉSTAMO hacia la barra de columnas y arrastramos la dimensión CARRERA hacia la barra de marcas; para mejorar el aspecto del reporte filtramos por año, mes y carrera, de manera que quedaría tal como muestra la figura:

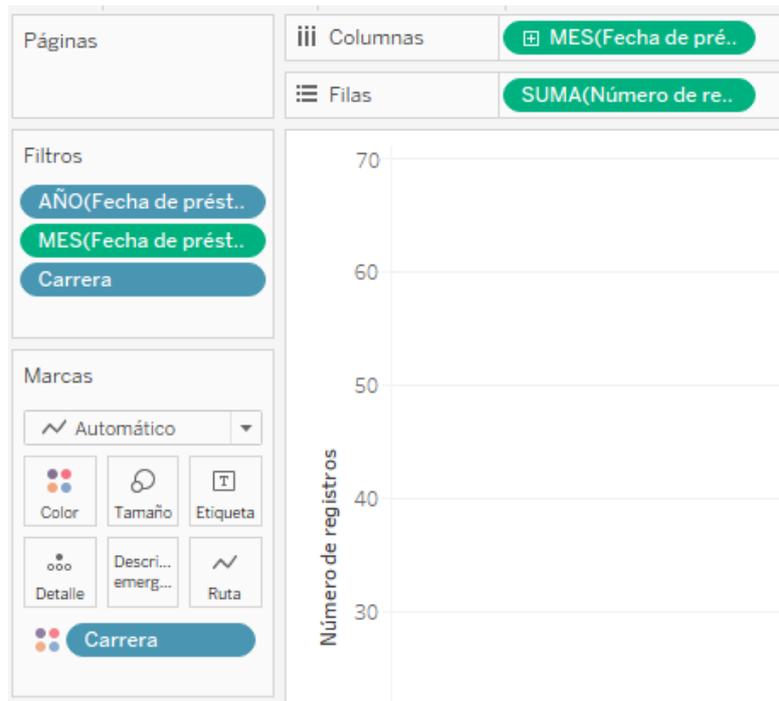


Figura 27. Vista de configuración de dimensiones y medidas.
Fuente: Elaboración propia.

Una vez realizada la configuración mencionada, nuestro reporte está listo para su visualización tal como se muestra a continuación:

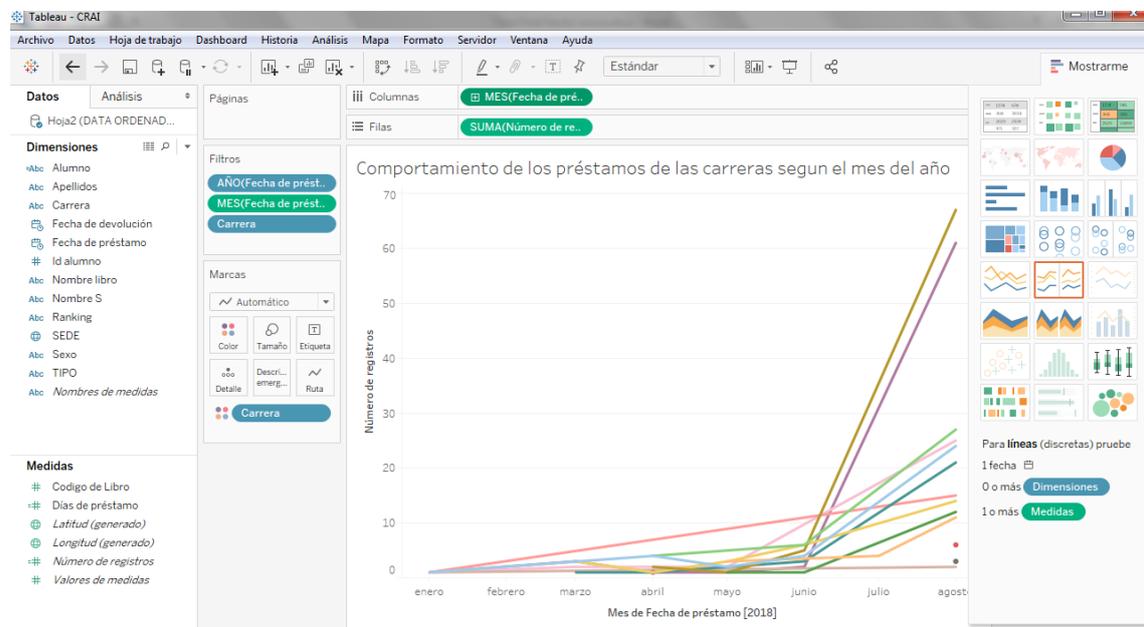


Figura 28. Elaboración del reporte de préstamos de las carreras según el mes del año.
Fuente: Elaboración propia.

2. Desarrollo del reporte de cantidad de préstamos total según cada mes del año

Para este reporte se arrastró la dimensión fecha de préstamo en la barra columnas desglosando en año y mes, también se arrastró la medida número de registros hacia la barra de filas así como el año de la fecha de préstamo hacia la barra de marcas, como se muestra a continuación:

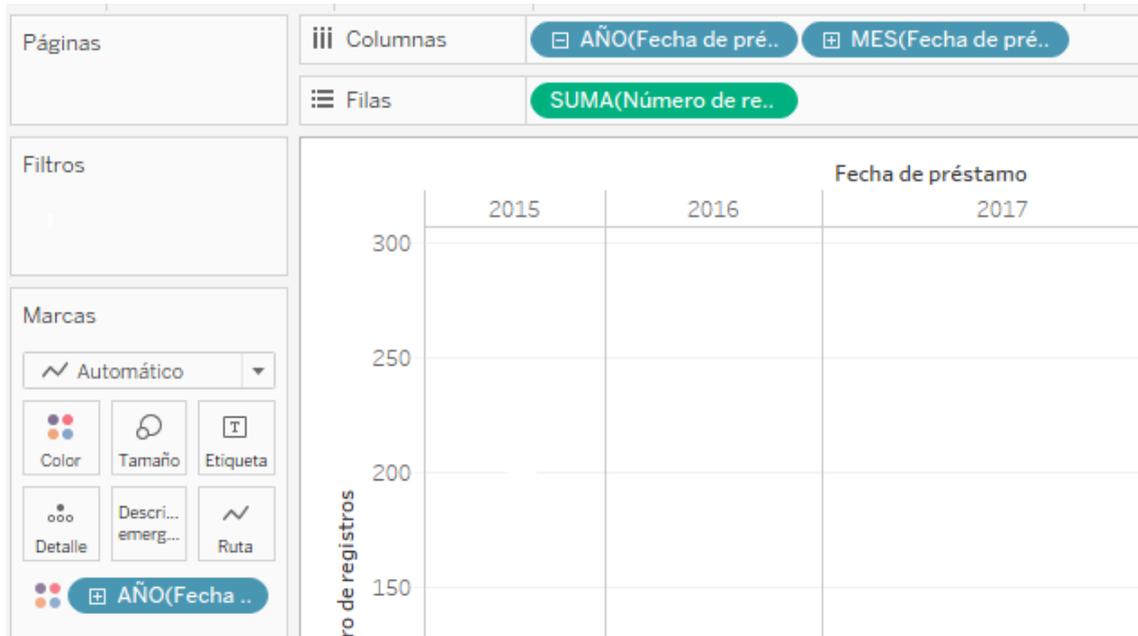


Figura 29. Configuración de cantidad de préstamos según total de cada mes del año.
Fuente: Elaboración propia.

Al realizar la configuración nuestro reporte estará listo para la visualización como lo muestra la siguiente figura:

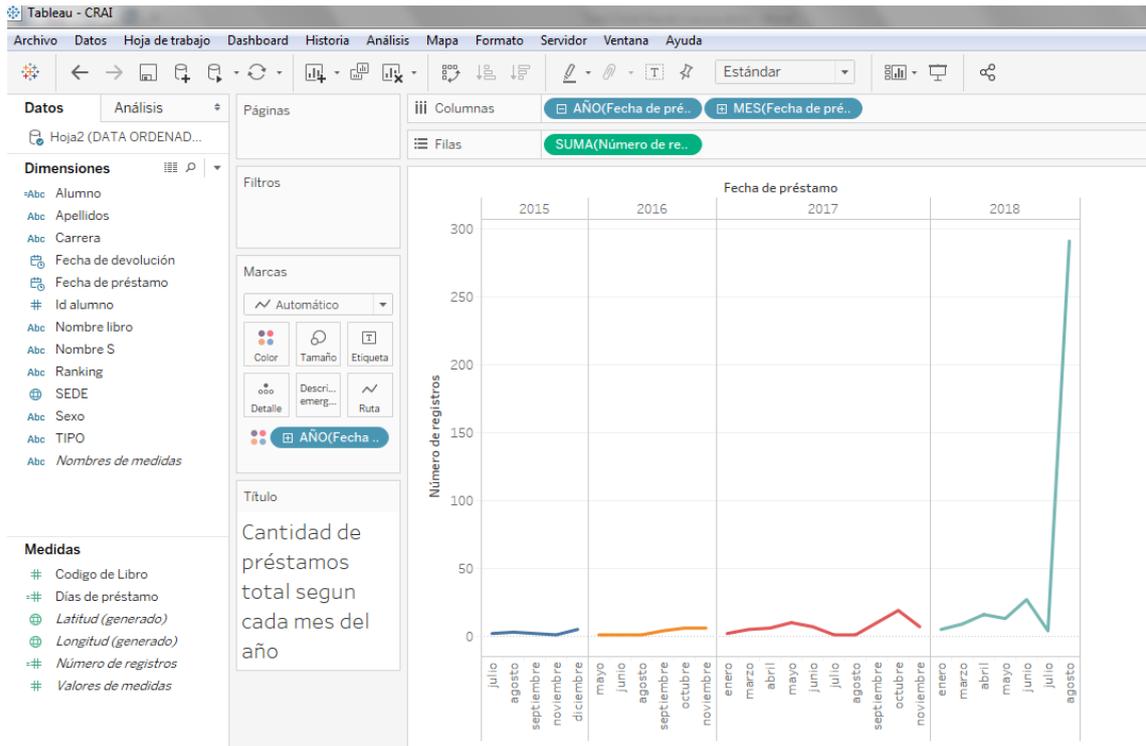


Figura 30. Elaboración del reporte de préstamos de las carreras según el mes del año.
Figura: Elaboración propia.

3. *Desarrollo del reporte de pronóstico de cantidad de préstamos total según cada mes del año*

Para este reporte se usó el reporte anterior añadiendo la opción pronóstico que provee Tableau en su panel de análisis y arrastrándolo hacia la barra de marcas como lo señalan la *Figura 31* y la *Figura 32*:



Figura 31. Panel de análisis de Tableau.
Fuente: Elaboración propia.

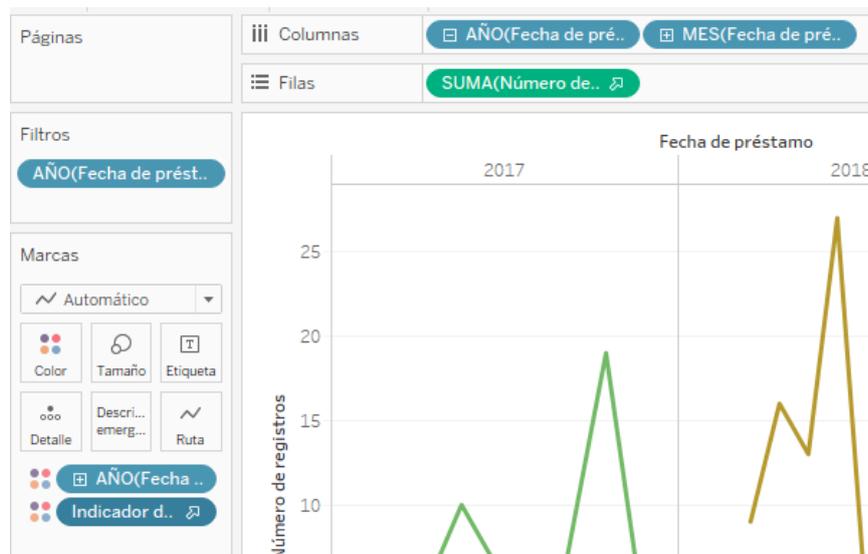


Figura 32. Configuración del pronóstico de cantidad de préstamos según el mes del año.
Fuente: Elaboración propia.

Al realizar la configuración nuestro reporte estará listo para la visualización como lo muestra la siguiente figura:

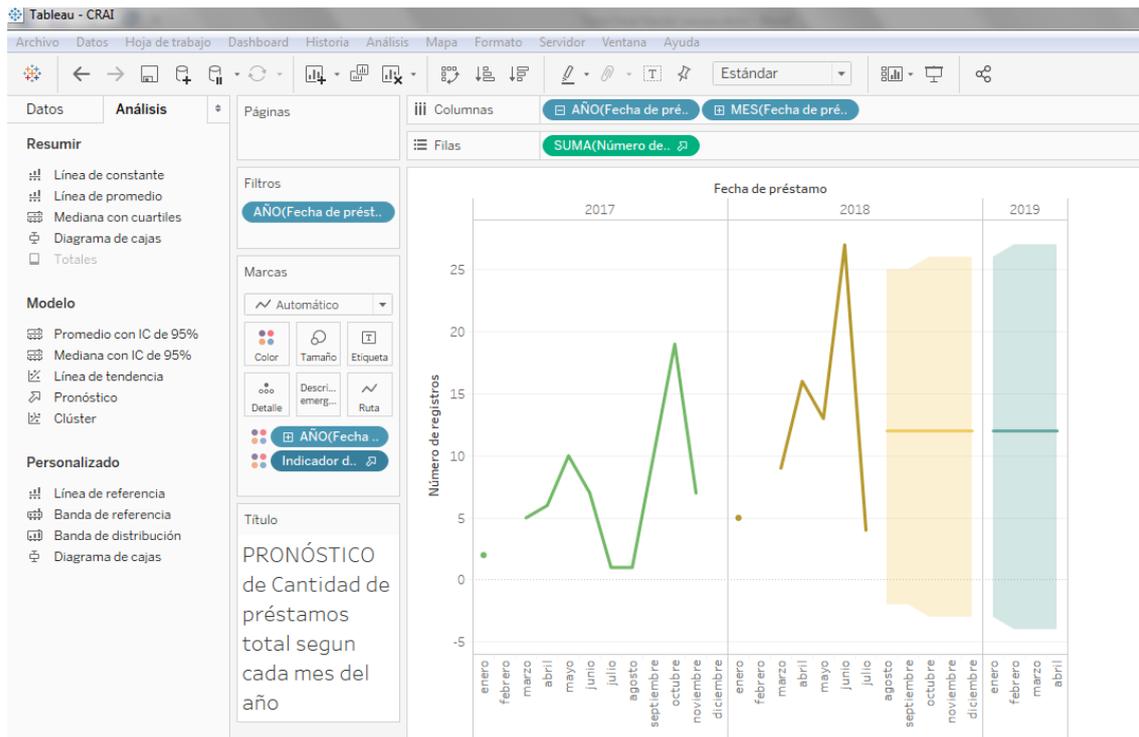


Figura 33. Elaboración del reporte de pronóstico de préstamos de las carreras según el mes del año.

Fuente: Elaboración propia.

4. Desarrollo del reporte de promedio de la cantidad de préstamos hechos según la hora del día y el año

Para este reporte se arrastró la dimensión fecha de préstamo en la barra columnas desglosando en año y hora, también se arrastró la medida número de registros hacia la barra de filas, así como el año de la fecha de préstamo hacia la barra de filtros, como se muestra a continuación:

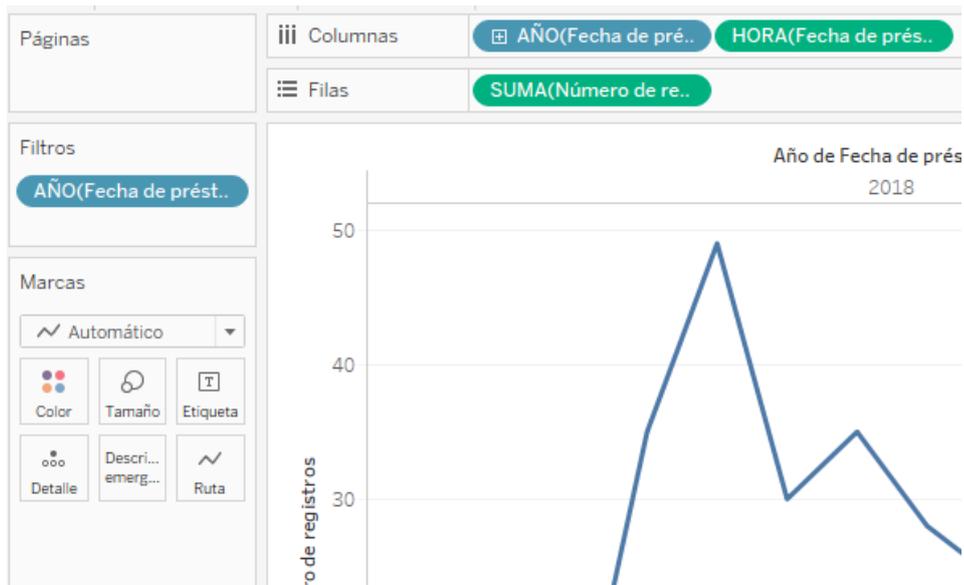


Figura 34. Configuración del reporte del promedio de cantidad de préstamos hechos según la hora del día y año.
Fuente: Elaboración propia.

Al realizar la configuración nuestro reporte estará listo para la visualización como lo muestra la siguiente figura:

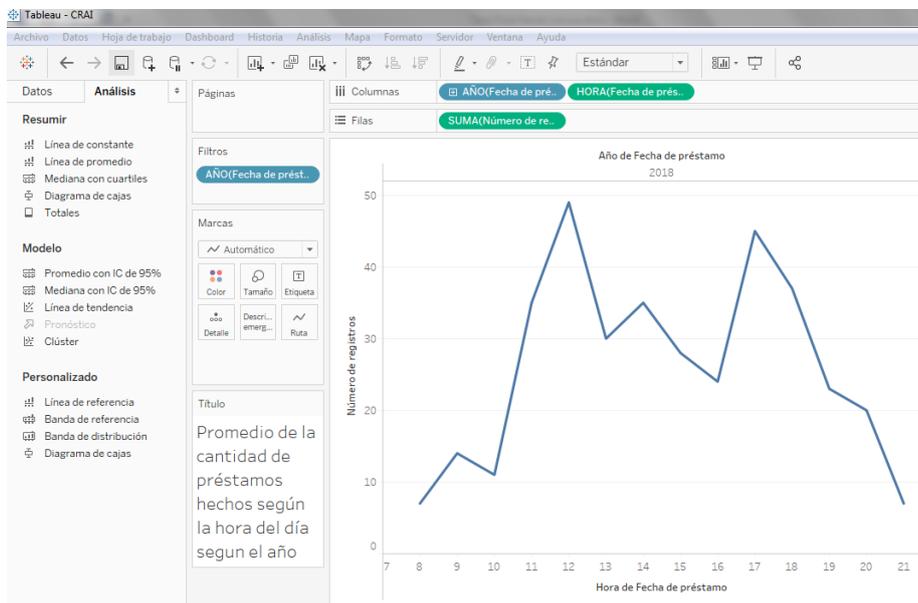


Figura 35. Elaboración del reporte del promedio de cantidad de préstamos hechos según la hora del día y año.
Fuente: Elaboración propia.

5. Desarrollo del reporte la cantidad de préstamo total de los estudiantes según su sexo

Para este reporte se arrastró la medida número de registros y la dimensión sexo hacia la barra de marcas, como se muestra a continuación:



Figura 36. Configuración del reporte de cantidad de préstamos total de los estudiantes según su sexo.

Fuente: Elaboración propia.

Otra configuración más que se realizó fue la opción de cálculo de tablas rápido que permitió el cálculo del porcentaje del total que se muestra en la Figura 37.



Figura 37. Configuración del porcentaje del total.

Fuente: Elaboración propia.

Al realizar la configuración nuestro reporte estará listo para la visualización como lo muestra la siguiente figura:

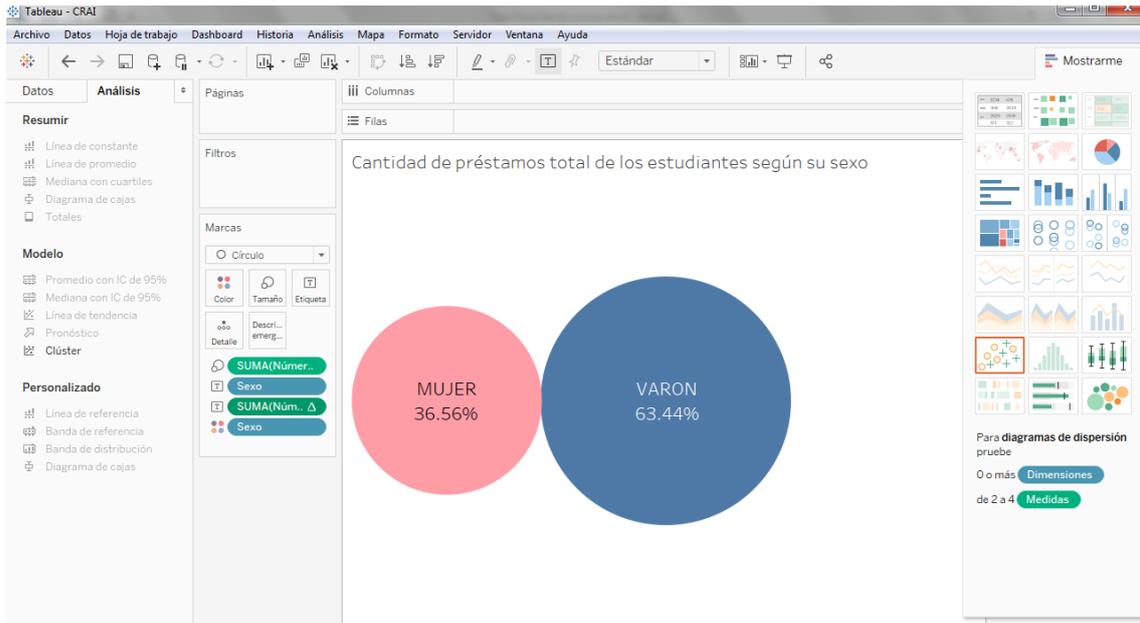


Figura 38. Elaboración del reporte de cantidad de préstamos total de los estudiantes según su sexo.
Fuente: Elaboración propia.

6. Desarrollo del reporte de duración de préstamos según la carrera

Para este reporte se arrastró la medida días de préstamo barra columnas, también se arrastró la dimensión carrera hacia la barra de filas y la barra de marcas, como se muestra a continuación:

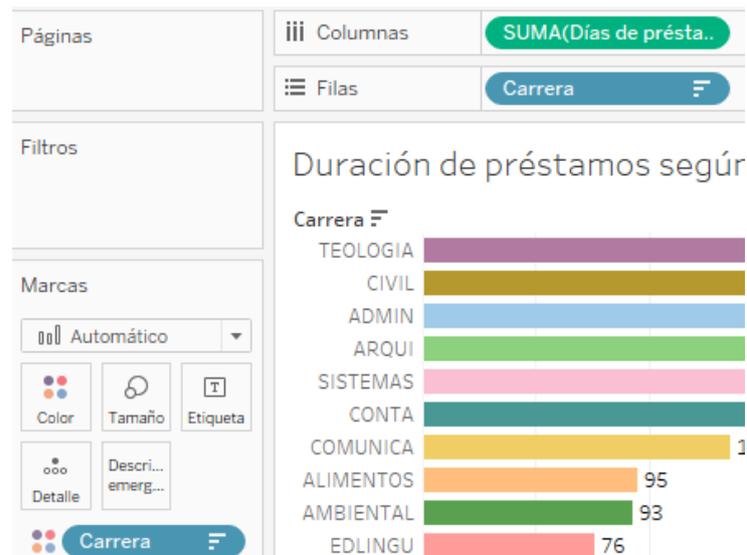


Figura 39. Configuración del reporte de duración de préstamos según la carrera.
Fuente: Elaboración propia.

Al realizar la configuración nuestro reporte estará listo para la visualización como lo muestra la siguiente figura:

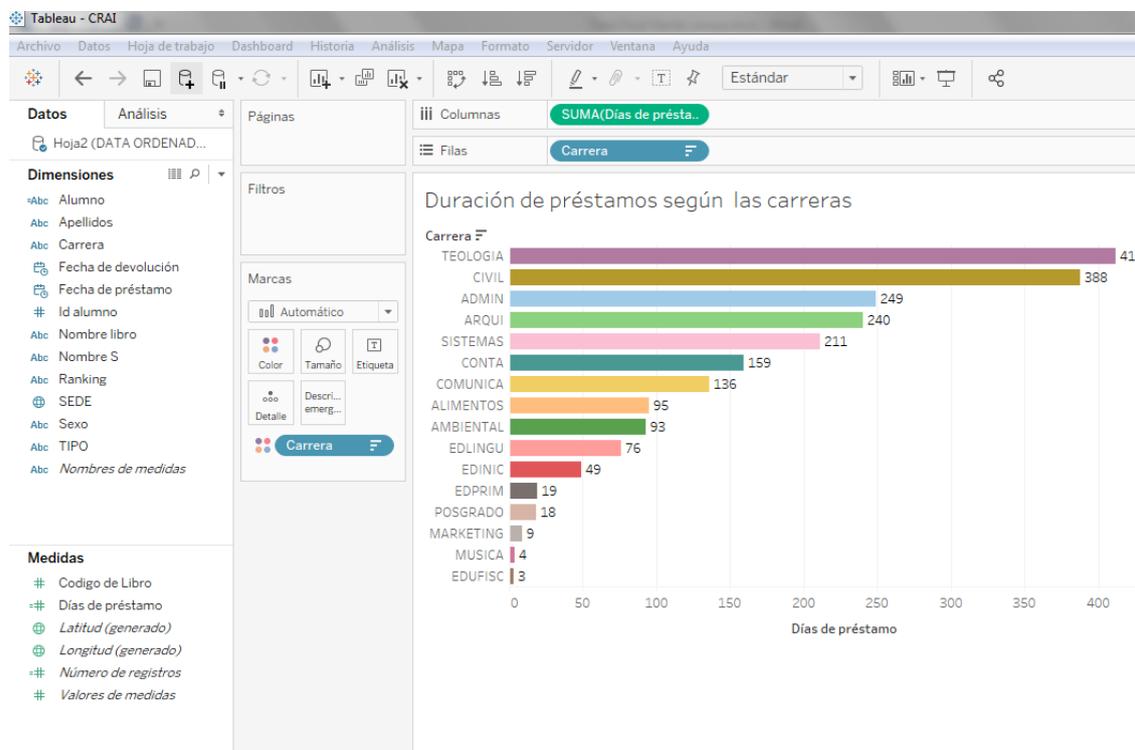


Figura 40. Elaboración del reporte de duración de préstamos según la carrera.
Fuente: Elaboración propia.

7. *Desarrollo del reporte de duración de préstamos de libros según la carrera y sexo*

Para este reporte se arrastró la dimensión carrera hacia la barra de columnas y la barra de filas se arrastró la medida días de préstamo, la dimensión sexo se arrastró a la barra de marcas, como se muestra a continuación:

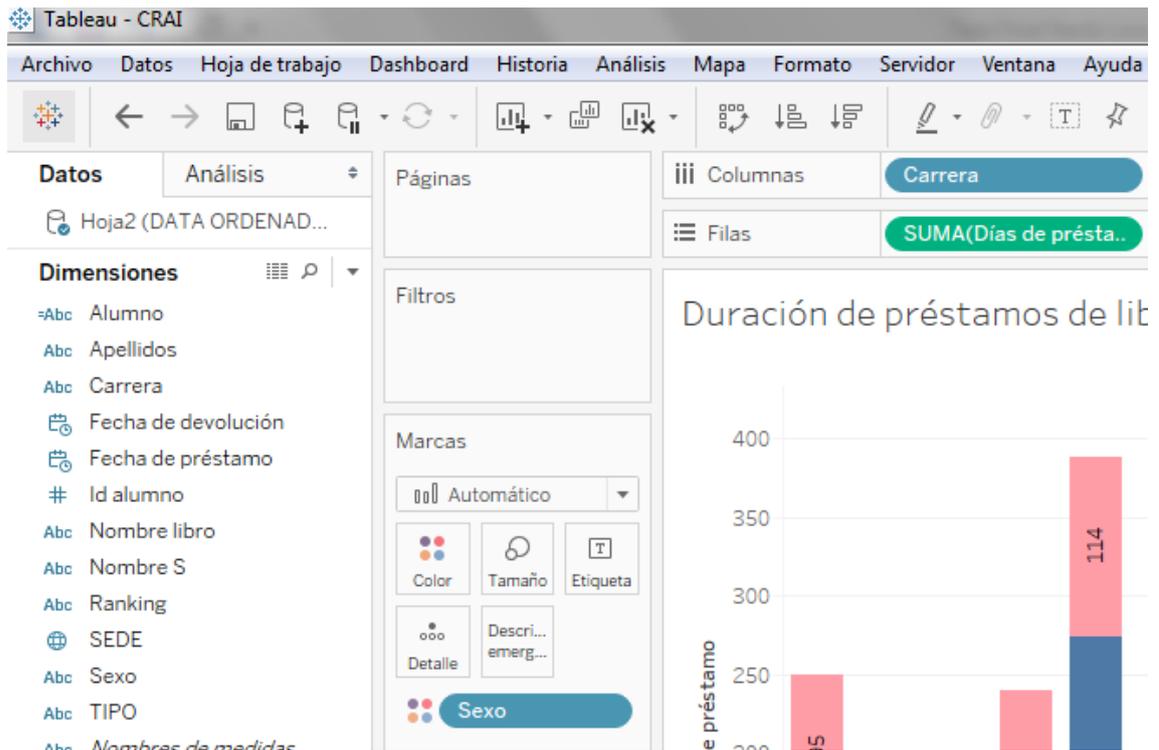


Figura 41. Configuración de préstamos de libros según carrera y sexo.
Fuente: Elaboración propia.

Al realizar la configuración nuestro reporte estará listo para la visualización como lo muestra la siguiente figura:

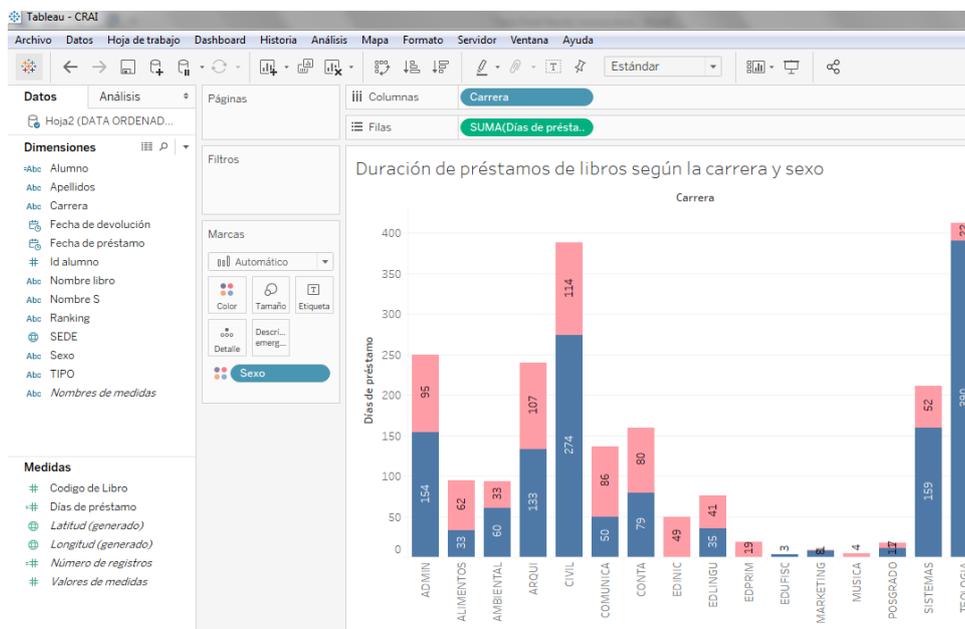


Figura 42. Elaboración de préstamos de libros según carrera y sexo.
Fuente: Elaboración propia.

8. *Desarrollo del reporte de Clustering (agrupamiento) de alumnos según la cantidad de días de préstamos de libros*

Para este reporte se arrastró la medida días de préstamo y la dimensión alumno hacia la barra de marcas; como se muestra a continuación:

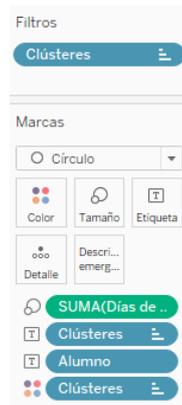


Figura 43. Configuración del reporte de clustering de alumnos según la cantidad de días de préstamo.

Fuente: Elaboración propia.

Dentro de esta configuración se arrastró también la opción de Clústeres que se encuentra en el panel de análisis de Tableau, como lo muestra la figura



Figura 44. Panel de análisis de Tableau con la opción de clúster.

Fuente: Elaboración propia.

Al realizar la configuración nuestro reporte estará listo para la visualización como lo muestra la siguiente figura:

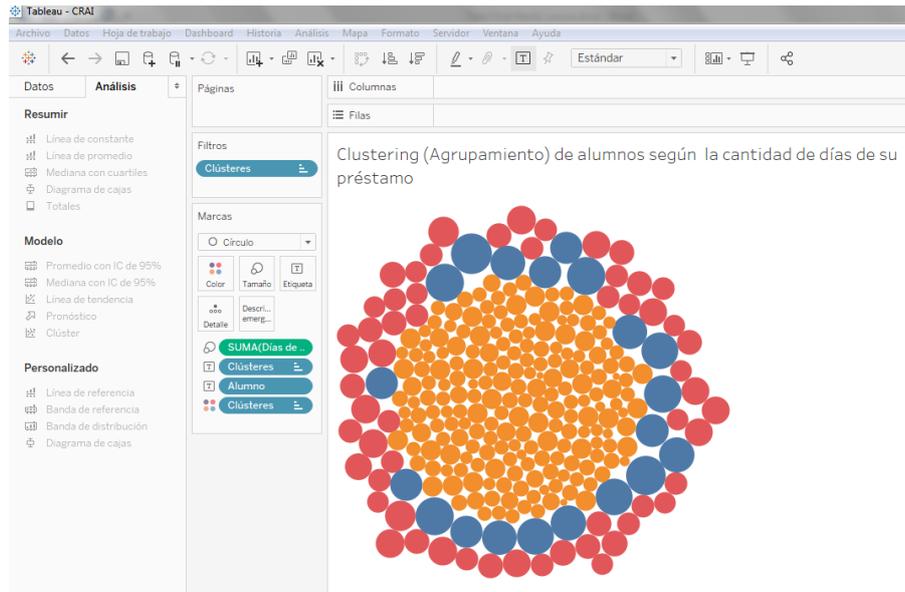


Figura 45. Elaboración del reporte de Clustering de alumnos según la cantidad de días de préstamo.

Fuente: Elaboración propia.

9. *Desarrollo del reporte del top 10 de los libros más recomendados.*

Para este reporte se usó la herramienta de software Spyder, en la cual se compiló el código que permite mostrar el top 10 de los libros más recomendados con el algoritmo KNN (algoritmo de vecino más cercano) el cual es un algoritmo de similitud que usa la distancia coseno.

```

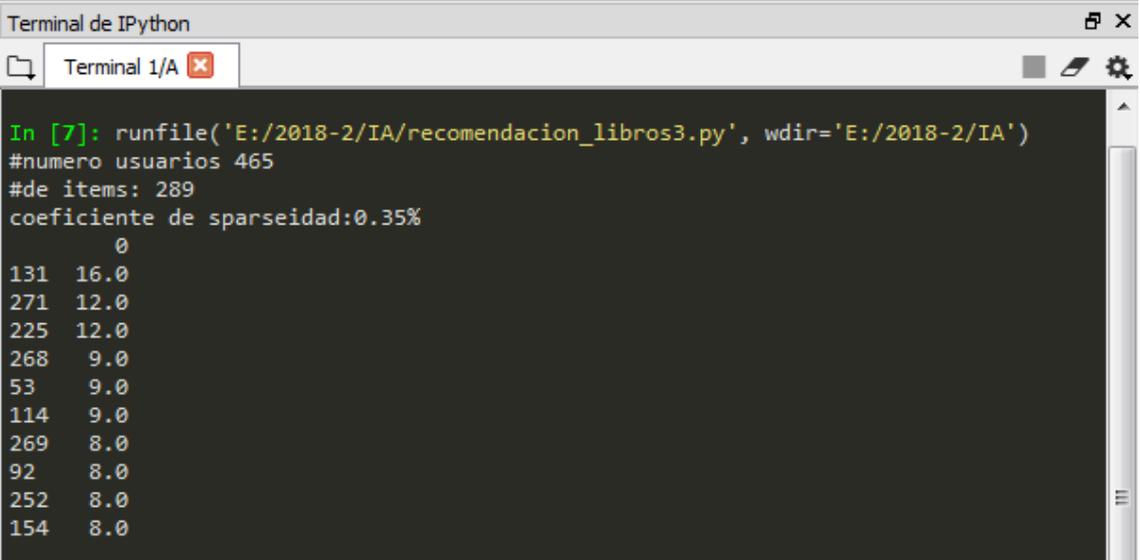
1 """
2 Created on Wed Sep 19 23:01:28 2018
3
4
5 @author: Narda
6 """
7
8 import numpy as np
9 import pandas as pd
10 import matplotlib.pyplot as plt
11 import seaborn as sb
12 import sklearn
13
14 #carga un dataframe
15 X= pd.read_csv("E:/2018-2/IA/data_final2prueba.csv",encoding="latin-1",sep=",").drop(
16     ["num"],
17     axis=1)
18
19 #cambiar el nombre de las columnas del dataset
20 X.columns=["Codigo_Libro","Codigo_alumno","ranking_sumatoria"]

```

Figura 46. Código que muestra el top de los 10 libros más recomendados.

Fuente: Elaboración propia.

Al ejecutar el código nos arrojó el siguiente resultado que se muestra en la siguiente figura:



```
Terminal de IPython
Terminal 1/A x

In [7]: runfile('E:/2018-2/IA/recomendacion_libros3.py', wdir='E:/2018-2/IA')
#numero usuarios 465
#de items: 289
coeficiente de sparseidad:0.35%
 0
131 16.0
271 12.0
225 12.0
268 9.0
53 9.0
114 9.0
269 8.0
92 8.0
252 8.0
154 8.0
```

Figura 47. Top 10 de libros más recomendados.
Fuente: Elaboración propia.

Estos resultados se interpretan en el capítulo v resultados y discusiones.

3.4.1.8. Implementación

Se implementó la plataforma BI con las herramientas seleccionadas y las especificaciones y requerimientos del usuario.

CAPÍTULO IV. RESULTADOS Y DISCUSIÓN

4.1. Resultados

4.1.1. Resultado 1

Uno de los logros alcanzados al implementar una plataforma de BI dentro del CRAI, fue reducir el tiempo de obtención de los reportes. Tal como lo mostró la *Tabla 1*, la demora en obtener reportes sin usar una plataforma BI es de 121 minutos; sin embargo, con Tableau el tiempo fue de 30 segundos, con lo cual se demuestra reducción del 99.58% del tiempo.

4.1.2. Resultado 2

Tableau es una gran herramienta que permitió mostrar el comportamiento de los préstamos de libros de las carreras, según el mes del año. En este caso, en la *Figura 48* mostró que las líneas de tendencia son mucho más crecientes en la escuela de teología (desde marzo) y la escuela de ingeniería civil a partir del mes de abril del 2018. Demostrando que éstas son las escuelas que más acceden al préstamo de literatura.

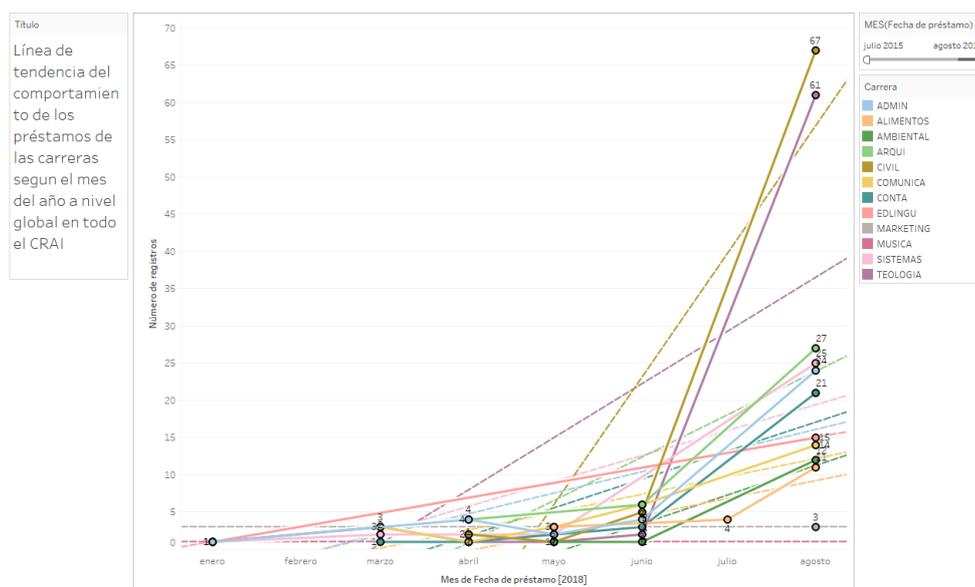


Figura 48. Línea de tendencia del comportamiento de préstamo de libros por mes. Fuente: Elaboración propia.

¿Porqué los estudiantes de ingeniería civil son los que más acceden al préstamo en el CRAI? Barranco (2012) en su investigación “Estrategias de lectura y escritura en Ingeniería Civil” afirma que el 84.7% de los docentes de la Pontificia Universidad Católica Madre y Maestra incluyen en su enseñanza la lectura. Dentro de ella sólo se consideran dos tipos de literatura: el 94.1% son libros de la asignatura y el 76.5% son libros de consulta de la asignatura. Esto indica que al igual que los docentes de ingeniería civil de la Pontificia Universidad Católica Madre y Maestra; los docentes de la escuela de ingeniería civil de la Universidad Peruana Unión usan literatura para sus asignaturas y que estén en la biblioteca, de manera que los estudiantes de ingeniería civil accedan al préstamo de éstos libros para sus consultas.

4.1.3. Resultado 3

Asimismo, la herramienta permitió visualizar los préstamos de libros según el mes, de manera que el Jefe del CRAI supo en qué mes del año hay más préstamos de libros, de modo que así en un futuro se pueda implementar un plan de contingencia ante alguna eventualidad de falta de libros y tomar acciones para que los alumnos devuelvan la literatura en su totalidad. Se observa que hay un incremento de préstamos durante los meses de julio a agosto.

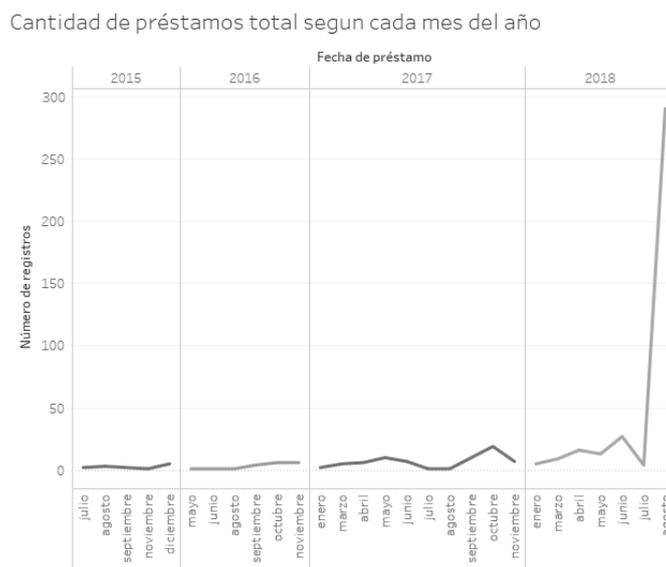


Figura 49. Cantidad de préstamos según el mes del año.
Fuente: Elaboración propia.

¿Por qué hubo un incremento de préstamos durante los meses de julio a agosto dentro del CRAI durante los últimos cuatro años? La Universidad Peruana Unión durante los

últimos años ha incrementado los diversos programas de estudios; uno de ellos es el ciclo regular de estudios que finaliza a fines del mes de junio, por lo cual los estudiantes de ciclo de regular acceden más al préstamo de libros durante el fin de ciclo. Otro de los programas de estudios que brinda la UPeU son los cursos extraprogramáticos durante los meses de vacaciones (julio y diciembre) en el cual los alumnos que desaprobaron alguna materia en la universidad tienen la oportunidad de llevarla nuevamente para que de esta manera puedan estar nivelados en su plan de estudios, estos también acceden a los recursos bibliográficos que ofrece el CRAI. Es por ello la razón del incremento de préstamos de literatura durante los meses de julio a agosto.

4.1.4. Resultado 4

El 63.5 % de los préstamos de literatura son a varones mientras el 36.5% son mujeres. Los estudiantes varones son quienes más acceden al préstamo de los libros, tal como lo muestra la *Figura 50*.

Cantidad de préstamos total de los estudiantes según su sexo

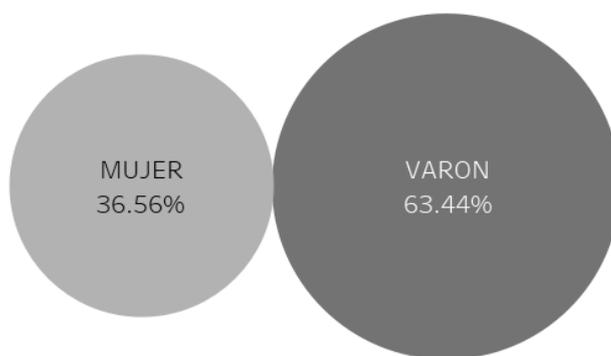


Figura 50. Cantidad de préstamos total según su sexo.

Fuente: Elaboración propia.

¿Por qué los varones acceden más al préstamo de la literatura que las mujeres en el CRAI? Gfk (2017) en su estudio “Frecuencia en la lectura de libros” indica que a nivel global las mujeres leen más que los hombres, 32% de ellas lee todos los días, frente a un 27% de ellos. En nuestra investigación los resultados reflejan una realidad distinta al estudio mencionado, la razón es que en la base de datos de matrícula de la Universidad Peruana Unión se muestra que los estudiantes de las escuelas de Teología e Ingeniería Civil son en

su mayoría varones; reafirmando el resultado N° 2, en que se observa que los estudiantes de las dos carreras mencionadas son quienes más acceden al préstamo de los libros.

4.1.5. Resultado 5

Según la *Tabla 13*, con el comportamiento de resultados y los gustos de los usuarios de biblioteca, el sistema de recomendación (nearest) recomienda al libro con código 131 que es: Introducción a la química ambiental; el libro corresponde a la categoría Ingenierías. Reafirmando el resultado obtenido con la herramienta Tableau, que mostraba que los estudiantes de Ingeniería civil son parte de quienes usan el préstamo de literatura, éste es el libro más recomendado ya que tiene la máxima puntuación en la tabla. El segundo libro más recomendado es de código 271, que es: Emociones e inteligencia social: Las claves para una alianza entre los sentimientos y la razón, el libro está correlacionado a los cursos de formación cristiana. Reafirmando el resultado obtenido con la herramienta Tableau, que mostraba que los estudiantes de Teología son parte de quienes usan el préstamo de literatura.

Tabla 13

Top de los 10 libros más recomendados

Código del libro	Puntuación
131	16.0
271	12.0
225	12.0
268	9.0
53	9.0
114	9.0
269	8.0
92	8.0
252	8.0
154	8.0

Fuente: Elaboración propia.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusión 1

Se concluye que la exhaustividad de la planificación del proyecto se logró implementar al 100%, lo cual el indicador mencionado es igual a 1. La actividad planificada de mostrar un ranking de los 10 libros más recomendados se realizó al 100% utilizando el algoritmo KNN. Los requerimientos, sustentados según los dashboards y resultados obtenidos se pudieron sustentar en su totalidad.

5.2. Conclusión 2

Se concluye Del 100% de data recibida de los años 2015 al 2018, se observó que el 57% era data no conforme para el proceso, por lo cual se tuvo que hacer una depuración de todo aquello que no era necesario para obtener una data óptima para la carga a Tableau, usando las funciones de Excel CONTAR.SI.CONJUNTO, TRUNCAR, SI, SUMAR.SI.CONJUNTO.

5.3. Conclusión 3

Se concluye que la plataforma de explotación de la información fue Tableau, la cual tiene una mejor capacidad de analizar y visualizar datos a comparación de Google Data Studio. A pesar que Tableau es una plataforma BI que requiere una licencia de pago, ofrece diversas opciones que ayudan a combinar los datos de múltiples fuentes en una sola vista, mientras que Google Data Studio no permite esta acción, quitándole potencia a pesar de ser libre de pago.

Como conclusión general. Se concluye que se implementó la plataforma BI con la herramienta Tableau con los datos de los años 2015, 2016,2017 y 2018 mostrando indicadores claves para la toma de decisiones.

Se recomienda que el CRAI pueda tener acceso remoto a la base de datos del SIRSI DYNIX SIMPHONY de manera que los datos puedan ser sincronizados al Tableau.

Se recomienda que esta solución mostrada en este proyecto sea implementada en bibliotecas de otras universidades de manera que se facilite la toma de decisiones beneficiando a los estudiantes de todo el Perú.

Se recomienda que tal como implementó una plataforma de BI en el CRAI, para la mejora de toma de decisiones, se pueda hacer lo mismo en otras áreas de la Universidad Peruana Unión, a fin de tener un datawarehouse de toda la universidad y mejorar la toma de decisiones dentro de ella.

INVESTIGACIONES FUTURAS:

La presente investigación fue hecha con datos de los años 2015, 2016, 2017 y 2018; se recomienda que esta investigación sea hecha con la data de CRAI sincronizada en línea para ver los cambios de los datos en tiempo real

REFERENCIAS BIBLIOGRÁFICAS

- Ad Tech & Analytics. (2017). El algoritmo K-NN y su importancia en el modelado de datos | Analítica web. Recuperado 23 de septiembre de 2018, de <https://www.analiticaweb.es/algoritmo-knn-modelado-datos/>
- Anselmo, M., & Espinoza, R. (2015). *Implementación De Un Datamart Para Toma De Decisiones De Recibos Distribuidos a Clientes Postpago De Telefónica Del Perú*.
- Artés, D. (2011). Proyecto fin de carrera 2008, 118-170.
- Barranco, A. (2012). Estrategias de lectura y escritura en Ingeniería Civil Reading and Writing Strategies in Civil Engineering, 9-15.
- Biblioteca Pública de Denver. (s. f.). Sistema de Clasificación Dewey | Biblioteca Pública de Denver. Recuperado 11 de septiembre de 2018, de <https://bibliotecadenver.org/sistema-de-clasificacion-dewey>
- Birkbeck, S. (2017). Kimball sigue siendo relevante? Recuperado 17 de septiembre de 2018, de <https://datametrics.co.nz/insights/article/is-kimball-still-relevant>
- Braga, D. (2018). Google Data Studio: La nueva herramienta para equipos de marketing. Recuperado 29 de agosto de 2018, de <https://marketingdecontenidos.com/google-data-studio/>
- Castro, A. (2017). ¿Qué es Machine Learning y para qué sirve? Recuperado 11 de septiembre de 2018, de <https://www.inbest.cloud/comunidad/qué-es-machine-learning-y-para-qué-sirve>
- Chand, S. (2019). What is Talend | Introduction to Talend ETL Tool | Edureka. Recuperado 10 de abril de 2019, de <https://www.edureka.co/blog/what-is-talend-tool/>
- Contel, B. (2008). *Desarrollo de una solución business intelligence en una empresa del sector de alimentación*. Universidad Politécnica Valencia. Recuperado de [https://riunet.upv.es/bitstream/handle/10251/9127/PFC Blanca Contel - Desarrollo de una](https://riunet.upv.es/bitstream/handle/10251/9127/PFC%20Blanca%20Contel%20-%20Desarrollo%20de%20una)

Solucion BI.pdf?sequence=1

Córdoba, C. (2015). *Spyder: Un IDE diseñado para científicos Contenidos*. Recuperado de <http://scipy.org/conf/2015/activity/download/attachment.file.91491e98edf5fe57.5370796465722e706466.pdf%3Ffilename=spyder.pdf>

Couprie, P. (s. f.). Similarity matrix. Recuperado 17 de septiembre de 2018, de <http://logiciels.pierrecouprie.fr/telechargement/EANALYSIS/EAnalysisHelp/pages/similarityMatrixDescription.html>

CRAI. (2016). Plan operativo anual POA. Juliaca. Recuperado de <https://crai.upeu.edu.pe/>

Curto, J. (2011). *Introducción al Business Intelligence*. Recuperado de <http://books.google.com/books?id=iU3RAXYQXMkC&pgis=1>

Dertiano, V. (2016). Comparativa entre Inmon y Kimball. Recuperado 2 de septiembre de 2018, de <https://blog.bi-geek.com/arquitectura-comparativa-inmon-y-kimball/>

Durán, A. G. (2017). ¿Qué es Pentaho Data Integrator (PDI)? Recuperado 29 de agosto de 2018, de <https://openwebinars.net/blog/que-es-pentaho-data-integrator-pdi/>

EcuRed. (s. f.). Sistema de Clasificación Decimal Dewey - EcuRed. Recuperado 30 de septiembre de 2018, de https://www.ecured.cu/Sistema_de_Clasificación_Decimal_Dewey

Escobar, J., & Gaete, S. (2015). *Desarrollode una solución de Inteligencia De Negocios para procesos mineros*. PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO. Recuperado de http://opac.pucv.cl/pucv_txt/txt-4500/UCD4959_01.pdf

Excelfull. (2018). ¿Qué es Excel? | Aprende Funciones y Utilidades en ExcelFull.com. Recuperado 29 de agosto de 2018, de <https://excelfull.com/que-es-excel/>

Fernández, B. (2018). Tableau: software de business intelligence, dashboard y big data | Clarcat. Recuperado 29 de agosto de 2018, de <https://www.clarcat.com/tableau-software/>

- Garcés, M. (2015). *Estudio comparativo de metodologías e implementación de alternativas business intelligence opensource vs. propietarias en entornos tradicionales; caso prototipo en las pymes en el sector agroindustrial*. Universidad de las Américas. Recuperado de <http://200.24.220.94/handle/33000/2660>
- Gartner, I. (2018). Business Intelligence - BI - Gartner Glosario de TI. Recuperado 2 de septiembre de 2018, de <https://www.gartner.com/it-glossary/business-intelligence-bi/>
- Gfk, E. G. (2017). *Frecuencia en la lectura de libros*.
- Goncalves, P. (2017). Data Integration - Kettle | Hitachi Vantara Community. Recuperado 29 de agosto de 2018, de <https://community.hitachivantara.com/docs/DOC-1009855-data-integration-kettle>
- González, L. (2018). 6 IDEs para Machine Learning con Python - Ligdi González. Recuperado 23 de septiembre de 2018, de <http://ligdigonzalez.com/ide-para-machine-learning-con-python/>
- Google. (2018). Le damos la bienvenida a Data Studio - Ayuda de Data Studio (Beta). Recuperado 29 de agosto de 2018, de <https://support.google.com/datastudio/answer/6283323?hl=es>
- Huichi, W. (2015). *Implementación de una plataforma de Business Intelligence para la toma de decisiones en un centro de salud*.
- Lima, J. (2015). *Inteligencia de negocios para la toma de decisiones del departamento de cartera de la Cooperativa Finander*. Recuperado de <http://186.3.45.37/bitstream/123456789/3303/1/TUASIS006-2014.pdf>
- MicroStrategy. (2018). Potente análisis de datos para todos. Recuperado 30 de agosto de 2018, de <https://www.microstrategy.com/es/product/analytics>
- Microsystem. (2018). TABLEAU LIDERA CUADRANTE MÁGICO GARTNER DE BI 2018 - Microsystem. Recuperado 10 de septiembre de 2018, de

<https://www.microsystem.cl/tableau-lidera-gartner-bi-2018/>

Moya, R. (s. f.). Sistemas de Recomendación basados en Filtrado Colaborativo (K-Vecinos) - Jarroba. Recuperado 14 de septiembre de 2018, de <https://jarroba.com/sistemas-de-recomendacion-basados-en-filtrado-colaborativo-k-vecinos/>

Moya, R. (2013). *Svd aplicado a sistemas de recomendación basados en filtrado colaborativo*. UNIVERSIDAD POLITÉCNICA DE MADRID. Recuperado de http://informatico.ricardomoya.es/docs/ProyectoFinDeMaster_V1.pdf

Moya, R. (2015). Pandas en Python, con ejemplos -Parte I- Introducción. Recuperado 12 de octubre de 2018, de <https://jarroba.com/pandas-python-ejemplos-parte-i-introduccion/>

Mundobiblio. (2017). Vigencia de la Tabla de notación interna de Luis F. Málaga. Recuperado 11 de septiembre de 2018, de <https://mundobiblio.wordpress.com/2017/03/28/vigencia-de-la-tabla-de-notacion-interna-de-luis-f-malaga/>

OBS. (2018). Microsoft Excel: Análisis del Software | OBS Business School. Recuperado 29 de agosto de 2018, de <https://www.obs-edu.com/int/blog-project-management/diagramas-de-gantt/microsoft-excel-analisis-del-software>

Palomino, Y., & Andrango, R. (2015). Implementación de un sistema de información ejecutiva utilizando Inteligencia de Negocios para la eficaz interpretación de Indicadores de Atención y Afiliación en el Seguro Integral de salud para la administración de la Red de Salud de Huarochirí, 112.

Peña, G. (2015). *DISEÑO DE UNA ARQUITECTURA DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA DE COMPRAS DE SEGUROS BOLÍVAR*. UNIVERSIDAD LIBRE DE COLOMBIA. <https://doi.org/10.1145/3132847.3132886>

PQS PERÚ. (2015). Software de datos: solo el 10% de empresas toma decisiones basadas en el análisis | Tecnología. Recuperado 20 de agosto de 2018, de <https://www.pqs.pe/tecnologia/software-de-datos-empresas-peruanas-decisiones-analisis>

- PREDIQT. (2018). Clientes Perú. Recuperado 26 de agosto de 2018, de <http://www.prediqtdata.com/5-clientes.html>
- PyData. (2016). PyData - Conference Mission. Recuperado 29 de agosto de 2018, de <https://pydata.org/mission.html>
- Robbins, S., & Coulter, M. (2014). *Administración*.
- RPP. (2017). Situación editorial: panorama del libro en el Perú | RPP Noticias. Recuperado 21 de agosto de 2018, de <https://rpp.pe/cultura/literatura/situacion-editorial-panorama-del-libro-en-el-peru-noticia-1045964>
- SAIMA Solutions. (2017). Soluciones QlikView - Partners de Qlik en Barcelona y Madrid. Recuperado 29 de agosto de 2018, de <https://saimasolutions.com/qlikview/>
- Santana, A., & Nieves, C. (2014). Objetos en R: Data Frames y Listas. Recuperado 12 de octubre de 2018, de http://www.dma.ulpgc.es/profesores/personal/stat/cursor4ULPGC/6g-Data_frames-Listas.html
- SirsiDyNix. (2017). SirsiDyNix Symphony | SirsiDyNix.com. Recuperado 30 de agosto de 2018, de <http://www.sirsidyNix.com/products/sirsidyNix-symphony>
- Tableau Software. (2017). ¿Qué es Tableau? | Tableau Software. Recuperado 29 de agosto de 2018, de <https://www.tableau.com/es-es/products/what-is-tableau>
- Valera, R. (1960). Hebreos 11 La fe Es, pues, la fe la certeza de - Bible Gateway. Recuperado 14 de octubre de 2018, de <https://www.biblegateway.com/passage/?search=Hebreos+11&version=RVR1960>
- Watkins, D. (2018). Getting started with Anaconda Python for data science | Opensource.com. Recuperado 29 de agosto de 2018, de <https://opensource.com/article/18/4/getting-started-anaconda-python>
- White, E. G. De. (1911). *Conflicto De Los Siglos. Conflicto de los Siglos*.

Zegarra, G. (2015). *Solución de inteligencia de negocios orientada a mejorar la toma de decisiones en las operaciones mineras de extracción y metalurgia de hochschild mining*. Recuperado de <http://www.repositorioacademico.usmp.edu.pe/handle/usmp/1827%0A>

ANEXOS

Anexo A. Entrevista

Jueves 01 de noviembre del 2018

Entrevista con el Jefe de Tecnologías de la Información del CRAI.

Nombre del entrevistado: Ing. Walter E. Luque Condori

1.- ¿Cuánto tiempo se demora en programar los reportes del Sistema de Biblioteca SirsiDynix Symphony Workflows?

Tener los resultados depende de la estabilidad de conexión a internet, el tiempo mínimo es de 2 minutos y el máximo oscila entre 5 minutos.

2.- ¿Esta demora es por el reporte de todas las escuelas o por cada escuela profesional?

Es por cada escuela profesional; por ejemplo, si se desea saber de la facultad de ciencias de la salud, ésta tiene dos escuelas: psicología y enfermería, y por cada una de ellas se tendría que emplear de 2 a 5 minutos.

3.- ¿En qué extensión de archivo es el resultado de las consultas generadas en el sistema?

En extensión XML, el cual es trabajado en Excel.

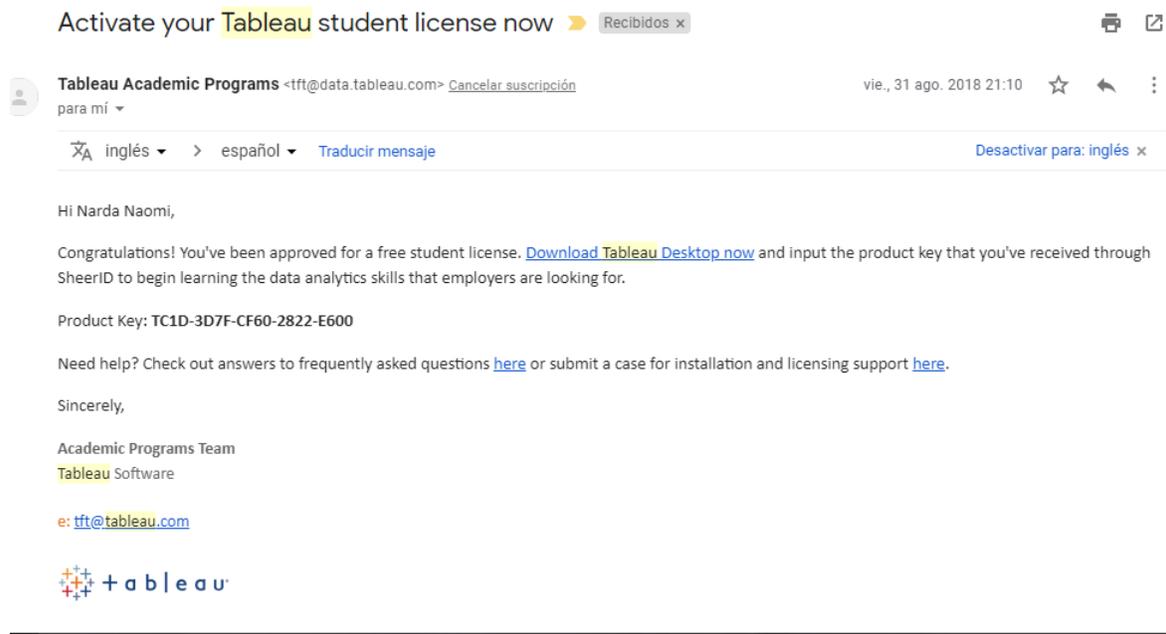
4.- ¿Cuánto tiempo se demora en generar las consultas en XML?

El tiempo mínimo es de 4 minutos y el máximo oscila entre 6 minutos.

Anexo B. Presupuesto de la investigación

Proyecto autofinanciado por el investigador. *En caso de Tableau proporciona una licencia académica gratuita al estudiante, ello con fines de investigación.

Imagen del correo recibido por Tableau donde otorga la licencia de estudiante:



Ítem	Descripción	Unidad	Cantidad	Precio S/	Parcial S/
01.01	SOFTWARE				
01.01.01	Tableau Professional Edition *	Mes	1	230.50	230.50
01.01.02	Google Data Studio	Mes	1	0.00	0.00
01.01.03	Python	Mes	1	0.00	0.00
01.01.04	Spyder	Mes	1	0.00	0.00
01.01.05	Kettle	Mes	1	0.00	0.00
	Subtotal				230.50
01.02	SERVICIOS				
01.02.01	Servicio de internet	Mes	3	30.00	90.00
	Subtotal				90.00
01.03	HARDWARE				
01.03.01	Computadora portátil	Horas	624	624.00	624.00
	Subtotal				624.00
01.04	RRHH				
01.04.01	Personal de ejecución del proyecto	Mes	3	1 000.00	3 000.00
	Consultoría de tesis	Mes		80.00	
	TOTAL				3714.00

Anexo C. Autorización de la investigación



Una Educación Avanzada



Juliaca, 10 de enero de 2019

Bachiller:

Narda Naomi Lanuza Bustamante

Bachiller de Facultad de Ingeniería y Arquitectura

Escuela Profesional de Ingeniería de Sistemas

Apreciada Bachiller.

Reciba un cordial saludo y deseamos éxitos en la labor que desempeña.

El motivo de la presente es para comunicarle que el Centro de Recursos para el Aprendizaje y la Investigación (CRAI) de la Universidad Peruana Unión Filial Juliaca ha aceptado la autorización y aprobación para que desarrolle su proyecto titulado "Implementación de una plataforma de Business Intelligence para la toma de decisiones en el Centro de Recursos para el Aprendizaje y la Investigación (CRAI) de la Universidad Peruana Unión."

Sin otro particular, me despido de usted.

Atentamente,




Mgster Julissa Torres Acurio
Directora del CRAI - UPeU FJ

Anexo D. Análisis Costo- Beneficio

Costo	Beneficio
Plataforma de BI Tableau Desktop	Optimiza el tiempo de toma de decisiones mediante las funciones de los dashboards.
Personal de ejecución del proyecto	Persona capacitada en la implementación de una plataforma BI que dio la solución según su expertise.
Computadora	Mejora la productividad, almacenamiento de datos.

Anexo E. Costo anual de la implementación de la plataforma BI

Descripción	Costo Mensual	Costo Anual
Licencia de Tableau Professional Edition	S/ 233.41	S/2800.92
Internet	S/ 121.95	S/1463.4
Total	S/355.36	S/4264.32

Anexo F. Documentación del algoritmo

```
1. # -*- coding: utf-8 -*-
2. """
3. Created on Wed Sep 19 23:01:28 2018
4.
5. @author: Narda
6. """
7. #El sistema de recomendación tiene como datos de entrada el préstamo
de libros por código de parte de los usuarios. Para ello se importó las
librerías:
8. #Numpy: Para las matrices a usar
9. #Pandas: Para el uso del dataframe
10. #Sklearn: Para el aprendizaje del modelo
11.
12. import numpy as np
13. import pandas as pd
14. import sklearn
15.
16. #A continuación se muestra la lectura del archivo csv de la data
17. X= pd.read_csv("E:/2018-2/IA/data_final2prueba.csv",encoding="latin-
1",sep=",").drop(
18.     ["num"],
19.     axis=1)
20.
21. #Se obtiene las columnas:codigolibro, codigo alumno y
ranking sumatoria
22. X.columns=["Codigo_Libro","Codigo_alumno","ranking_sumatoria"]
23.
24. # Se obtiene cuantos usuarios hay en la base de datos
25. n_users=X.Codigo_alumno.unique().shape[0]
26. print("#numero usuarios",n_users)
27. #Se obtiene cuantos libros hay en la BD
28. n_ItemId=X.Codigo_Libro.unique().shape[0]
29. print("#de items:",n_ItemId)
30.
31. #Testing de la data y entrenamiento del modelo
32. ratings=np.zeros((n_users,n_ItemId))
33. for row in X.itertuples():
34.     ratings [row[2]-1,row[1]-1]=row[3]
35.
36.
37. #Se obtiene el indicador de sparseidad, el cual permite conocer las
distancias entre las preferencias de un usuario y otro, usando la lógica
de KNN vecinos, cuanto menor sea la distancia es mejor.
38. sparsity=float(len(ratings.nonzero()[0]))
39. sparsity/=(ratings.shape[0]*ratings.shape[1])
40. sparsity*=100
41. print("coeficiente de sparseidad:{:4.2f}%".format(sparsity))
42.
43. #conjunto de entrenamiento de aprendizaje automático supervisado
usando la validación cruzada
44. from sklearn.cross_validation import train_test_split
45. ratings_train, ratings_test=train_test_split(ratings,test_size=0.3,ran
dom_state=42)
```

```

46.
47. #definición del error medio cuadrático del modelo
48. from sklearn.metrics import mean_squared_error
49. #Usando el algoritmo de la distancia coseno para la matriz de
similaridad (descripción de usuarios respecto a las preferencias) se coloca
de manera numerica las preferencias que existen
50. sim_matriz=1-sklearn.metrics.pairwise.cosine_distances(ratings_train)
51. #Cada columna es un usuario y detro de ella tenemos distintas
preferencias, se observa cuan similares son las preferencias entre usuarios
52. users_prediction=sim_matriz.dot(ratings_train)/np.array([np.abs(sim_ma
triz).sum(axis=1)]).T
53.
54. #algoritmo de libros más populares, user_prediction matriz...numpy
55. a = np.shape(1682) #matriz de 1 columna por 1682 filas que son los
libros sumados de la ponderacion
56. a=users_prediction.sum(axis=0)# la suma del numpy es por columnas
57. populares_items=pd.DataFrame(a) #convierto el numpy a dataframe
58.
59. #cero en del sort es el nombre de la columna que se hara top
60. top_dataframe=populares_items.sort_values(0,ascending=False).head(10)
61. #reporte de las 10 libros más valorados imprime los libros mas
populares

print(top_dataframe)

```