

## Guía3. Informe final Proyecto APT Asignatura Capstone

### 1. Informe final Proyecto APT

El objetivo de este informe es que describas los aspectos más relevantes de tu Proyecto APT. Es importante que fundamentes las decisiones que tuviste que tomar a lo largo del proceso.

A continuación, encontrarás distintos campos que deberás completar con la información solicitada, los que dan cuenta del resumen de tu proyecto APT y sus principales resultados.

Nombre del proyecto	Sistema y Gestión Middleware ETL
Área (s) de desempeño(s)	Automatización, Ingeniería de datos y Desarrollo de software
Competencias	<p>Ofrecer propuestas de solución según los requerimientos de la organización.</p> <p>Integración y correcta administración de los distintos ambientes informáticos, logrando una correcta operatividad por parte de la empresa.</p> <p>Desarrollo de la solución utilizando software de alto calibre, escalable y mantenible. asegurando el logro de los objetivos.</p> <p>Implementar soluciones sistémicas integrales para automatizar y optimizar procesos de negocio de acuerdo a las necesidades de la organización.</p> <p>Resolver las vulnerabilidades sistémicas para asegurar que el software construido cumple las normas de seguridad exigidas por la industria.</p> <p>Programar consultas o rutinas para manipular información de una base de datos de acuerdo a los requerimientos de la organización.</p>


























<p>1. Relevancia del proyecto APT</p>	<p>El proyecto busca solucionar la problemática de la fragmentación de datos en Recursos Humanos (RR.HH.) dentro de la empresa Cencosud, una empresa con operaciones a nivel internacional. La necesidad de acceder a información unificada de todos sus empleados en diferentes países se ve obstaculizada por la dispersión de los datos de RR.HH. en sistemas y bases de datos con estructuras y enfoques distintos, dificultando la creación de una visión global y coherente de la fuerza laboral.</p> <p>Esta fragmentación dificulta la toma de decisiones estratégicas basadas en datos a nivel global, la gestión del talento en diferentes regiones y la creación de programas de desarrollo efectivos que se adapten a las necesidades específicas de cada país. Esta problemática es crucial para el campo laboral de la gestión de datos y la analítica, ya que la necesidad de integrar y analizar datos de diferentes fuentes es una realidad en la mayoría de las empresas multinacionales.</p> <p>El proyecto propone la implementación de un "middleware" que actúe como puente entre las diferentes fuentes de datos a nivel internacional. Este middleware centralizará la información en un modelo unificado de entidades relacionadas entre sí, permitiendo:</p> <ul style="list-style-type: none"><li>● <b>Automatizar el proceso ETL:</b> Se implementará un proceso automatizado de extracción, transformación y carga de datos (ETL) para integrar la información de las diferentes fuentes, creando una base de datos centralizada y consistente a nivel global.</li><li>● <b>Integración de Clusters de Machine Learning:</b> El análisis de datos se verá enriquecido mediante la incorporación de algoritmos de aprendizaje automático, lo que permitirá obtener insights más profundos y generar predicciones a nivel global y regional.</li><li>● <b>Accesibilidad:</b> Los datos procesados podrán ser consumidos por diferentes plataformas, como Power BI, la plataforma web interna de la empresa, el equipo de científicos de datos, el departamento de inteligencia de negocio y RR.HH., tanto a nivel central como en las diferentes regiones.</li></ul> <p>Con esto aportamos una solución real al contexto laboral de Cencosud a nivel internacional, proporcionando una herramienta para integrar y analizar datos de RR.HH. de manera eficiente y accesible. Este sistema de datos centralizado y unificado, basado en un modelo de entidades interrelacionadas, facilitará la gestión de los diferentes sectores que requieren acceder a información de RR.HH. a nivel global, permitiendo optimizar la gestión del talento, mejorar la toma de decisiones estratégicas y fortalecer el desarrollo de la empresa a nivel internacional.</p>
---------------------------------------	---

2. Objetivos	<p>Como objetivos principales, se busca proporcionar un software autogestionable, escalable y mantenible en el tiempo. Para esto, desglosamos los objetivos específicos para este proyecto.</p> <ul style="list-style-type: none"> <li>• Proporcionar un sistema agnóstico a diferentes fuentes de datos.</li> <li>• Generar Gráficos proporcionales con gran valor estadístico.</li> <li>• Procesar tareas automáticas frecuentemente.</li> <li>• Sistema autogestionable.</li> <li>• Sistema con salida a escalabilidad.</li> <li>• Datos con gran valor estadístico y con salida a crecimiento constante.</li> </ul>
3. Metodología	<p>Actualmente contamos con metodología ágil, para este proyecto, específicamente SCRUM. Donde contaremos con una serie de factores positivos para la realización y la plenitud constante de este proyecto.</p> <p>Por un lado contaremos con los requerimientos principales, y levantamiento de la propuesta a realizar, esto nos ayudará a tener mejor conocimiento de las funcionalidades que debe cumplir el proyecto, evitando fallas de conceptos por parte del equipo SCRUM, esta fase ya fue realizada por parte de nuestro miembro con la función de Product Owner, donde realizó todo el proceso de gestión y comprensión inicial con nuestros stakeholders.</p> <p>Luego, una vez definido nuestros requerimientos y los estándares a cumplir, comenzaremos con los primeros SPRINTS, donde se recalcó el utilizar las historia de usuario como tareas a resolver. Estos sprints van por fase y se dió como máximo, 3 semanas para completar todas las tareas o historias de usuario dentro de un SPRINT. Para llevar la fase de los sprints por un buen camino, realizaremos reuniones diarias, o más conocidas como Daily Meeting, donde utilizaremos estas reuniones para dar a conocer nuestros avances o retrasos al estar resolviendo una tarea definida.</p> <p>Estos básicos conceptos de metodologías nos bastará para llevar por un buen camino el proyecto que se está realizando, además nos ayuda a mantener los requerimientos a flote ante cualquier cambio repentino, donde aprovecharemos los Sprints para dar a conocer nuevos requerimientos que surgieron, o conceptos de estos.</p>
4. Desarrollo	<p>Actualmente contamos con una serie de objetivos o requerimientos a utilizar, divididos por historias de usuarios. Donde utilizaremos los sprints como etapas en el desarrollo del proyecto. Al día de hoy, contamos con un mvp preliminar donde este cuenta con las funcionalidades principales del proyecto, el cual abordamos todo el proceso ETL, estadística descriptiva, y gestión de datos.</p> <p>Las dificultades al abordar este proyecto, es el tema de los datos generados artificialmente con lógica muy específica, ya que este</p>

proyecto cuenta con una solución enfocada al sector de recursos humanos de la empresa Cencosud, pero al ser datos sensibles y al estar protegido por regimens, se dificulta la utilización de estos para abordar el proyecto APT. A pesar de este problema muy importante, supimos resolverlo, y actualmente contamos con una fase muy avanzada del proyecto, el cual cumple con las expectativas impuestas inicialmente.

En un inicio, presentamos la solución principal, el cual aborda la problemática presentada al inicio de este documento. Pero como se recalcó, este desarrollo tuvo que realizarse con un enfoque de datos diferente, pero seguir con la idea de un desarrollo autogestionado, y sirviendo como intermediario entre varios sistemas de datos.

Principales pipelines de datos:

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
 dag_country_etl	arturo		None	2024-10-21, 02:24:00				
 dag_employee_etl	arturo		None	2024-10-20, 04:26:59				
 dag_hr_report_etl	arturo		None	2024-10-20, 05:34:07				
 dag_invoice_etl	arturo		None	2024-10-20, 00:01:32				
 dag_store_etl	arturo		None	2024-10-20, 02:49:30				

Bases de datos pobladas exitosamente con procesos ETL:

public	auth_group	table	postgres
public	auth_group_permissions	table	postgres
public	auth_permission	table	postgres
public	auth_user	table	postgres
public	auth_user_groups	table	postgres
public	auth_user_user_permissions	table	postgres
public	country	table	postgres
public	customer	table	postgres
public	django_admin_log	table	postgres
public	django_content_type	table	postgres
public	django_migrations	table	postgres
public	django_session	table	postgres
public	employee	table	postgres
public	hr_report	table	postgres
public	invoice	table	postgres
public	store	table	postgres

5. Evidencias

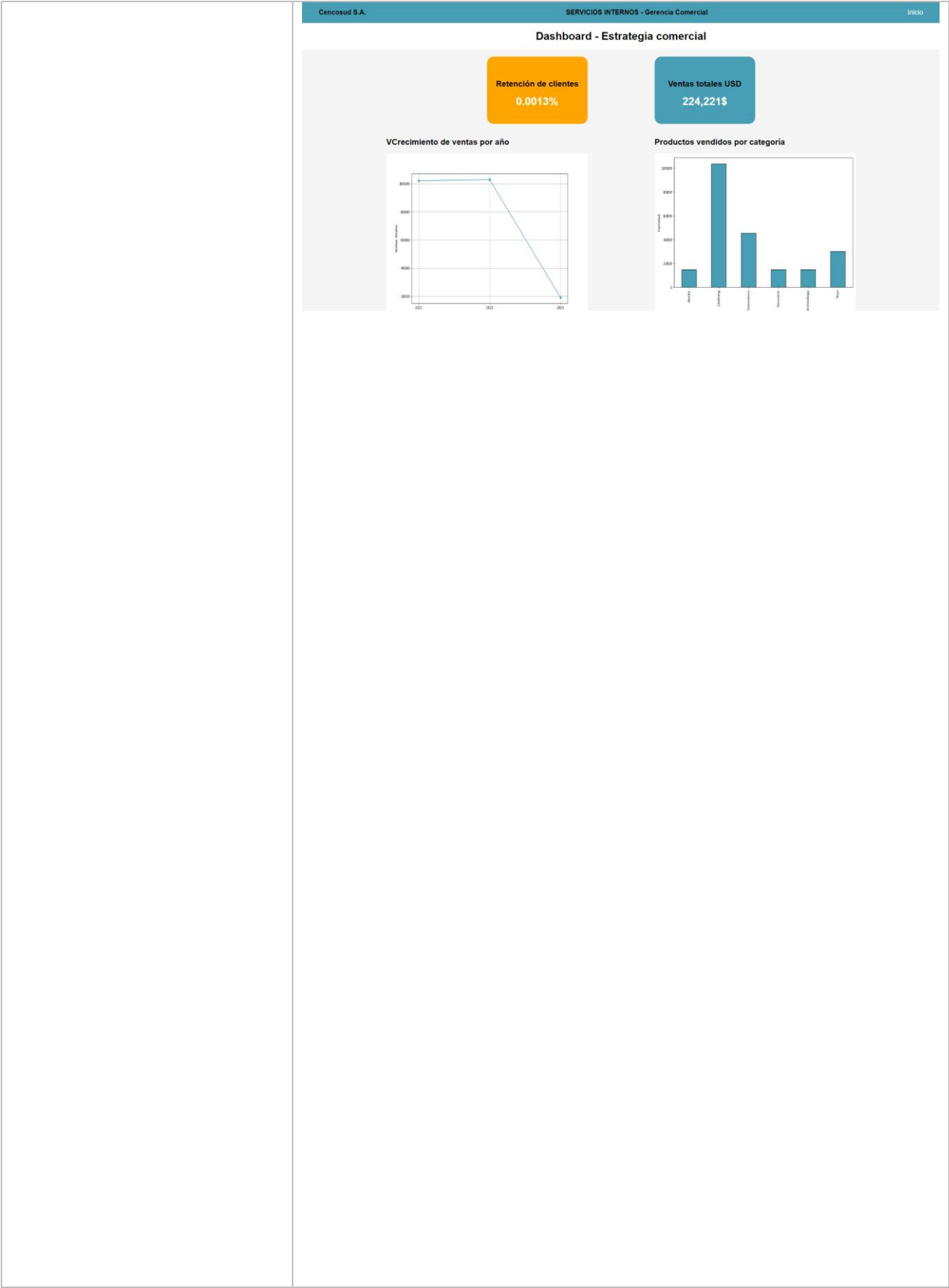
id	gender	age	name	last_name	email	nationality
C241288	Female	28	Georgina	Fernandez	Geo.Fernandez@gmail.com	Brasil
C266599	Male	20	Alberto	Vazquez	Alb.Vazquez@gmail.com	Colombia
C189076	Female	53	Isabel	Perez	Isa.Perez@gmail.com	Estados Unidos
C657758	Female	28	Laura	Jimenez	Lau.Jimenez@gmail.com	Peru
C151197	Female	49	Georgina	Ruiz	Geo.Ruiz@gmail.com	Mexico
C176086	Female	32	Georgina	Perez	Geo.Perez@gmail.com	Mexico
C159642	Male	69	Williams	Molina	Wil.Molina@gmail.com	Argentina
C283361	Female	60	Ana	Ortega	Ana.Ortega@gmail.com	Colombia
C191708	Female	29	Fernanda	Diaz	Fer.Diaz@gmail.com	Argentina
C225330	Female	67	Laura	Molina	Lau.Molina@gmail.com	Colombia
C312861	Male	25	Edgardo	Fernandez	Edg.Fernandez@gmail.com	Chile
C555402	Female	67	Javiera	Gonzalez	Jav.Gonzalez@gmail.com	Colombia
C300786	Male	65	Luis	Castro	Lui.Castro@gmail.com	Mexico
C218149	Female	46	Carmen	Fernandez	Car.Fernandez@gmail.com	Peru
C196845	Male	24	Juan	Ortega	Jua.Ortega@gmail.com	Espana
C220180	Male	23	Pedro	Sanchez	Ped.Sanchez@gmail.com	Australia
C322947	Male	52	Williams	Ramirez	Wil.Ramirez@gmail.com	Mexico
C313348	Male	44	Williams	Hernandez	Wil.Hernandez@gmail.com	Mexico
C204553	Female	42	Lucia	Rodriguez	Luc.Rodriguez@gmail.com	Argentina
C285161	Male	51	Pedro	Sanchez	Ped.Sanchez@gmail.com	Mexico
C289625	Female	25	Lucia	Ruiz	Luc.Ruiz@gmail.com	Peru
C251229	Male	29	Edgardo	Molina	Edg.Molina@gmail.com	Estados Unidos
C159164	Female	66	Lucia	Rodriguez	Luc.Rodriguez@gmail.com	Mexico
C501658	Male	23	Benjamin	Hernandez	Ben.Hernandez@gmail.com	Argentina
C232624	Male	51	Benjamin	Sanchez	Ben.Sanchez@gmail.com	Chile
C245456	Male	43	Benjamin	Rodriguez	Ben.Rodriguez@gmail.com	Australia
C450287	Female	59	Javiera	Martinez	Jav.Martinez@gmail.com	Argentina
C204279	Male	54	Williams	Gonzalez	Wil.Gonzalez@gmail.com	Chile
C452806	Male	24	Juan	Diaz	Jua.Diaz@gmail.com	Mexico
C155059	Male	67	Edgardo	Garcia	Edg.Garcia@gmail.com	Argentina
C306662	Female	48	Carmen	Moreno	Car.Moreno@gmail.com	Colombia
C134449	Male	46	Carlos	Molina	Car.Molina@gmail.com	Brasil
C317818	Female	19	Georgina	Lopez	Geo.Lopez@gmail.com	Brasil
C241642	Female	43	Laura	Castro	Lau.Castro@gmail.com	Brasil
C126436	Male	18	Alberto	Martinez	Alb.Martinez@gmail.com	Mexico
C653385	Female	22	Fernanda	Molina	Fer.Molina@gmail.com	Estados Unidos
C227070	Female	61	Ana	Vazquez	Ana.Vazquez@gmail.com	Peru
C317478	Female	41	Carmen	Perez	Car.Perez@gmail.com	Argentina

## Guía Estudiante – Informe Final Proyecto APT Fase 2

employee_id	first_name	last_name	education	age	email	gender	store_id
74430	Estefania	Rodriguez	Bachelor's	31	Est.Rodriguez@gmail.com	Female	4
38562	Georgina	Fernandez	Bachelor's	31	Geo.Fernandez@gmail.com	Female	2
46232	Anibal	Ramirez	Bachelor's	36	Ani.Ramirez@gmail.com	Male	7
54542	Jorge	Hernandez	Bachelor's	33	Jor.Hernandez@gmail.com	Male	2
66174	Anibal	Ortega	Master's & above	51	Ani.Ortega@gmail.com	Male	7
30963	Anastacia	Perez	Master's & above	34	Ana.Perez@gmail.com	Female	4
12737	Edgardo	Hernandez	Bachelor's	31	Edg.Hernandez@gmail.com	Male	7
31582	Lucia	Torres	Bachelor's	49	Luc.Torres@gmail.com	Female	7
72735	Carlos	Martinez	Master's & above	37	Car.Martinez@gmail.com	Male	9
51498	Williams	Perez	Master's & above	41	Wil.Perez@gmail.com	Male	4
72407	Benjamin	Garcia	Master's & above	34	Ben.Garcia@gmail.com	Male	2
65237	Anibal	Ortega	Bachelor's	43	Ani.Ortega@gmail.com	Male	2
77460	Ana	Sanchez	Bachelor's	35	Ana.Sanchez@gmail.com	Female	2
36834	Edgardo	Garcia	Master's & above	30	Edg.Garcia@gmail.com	Male	8
7935	Fernanda	Martinez	Bachelor's	52	Fer.Martinez@gmail.com	Female	2
49940	Luis	Sanchez	Bachelor's	26	Lui.Sanchez@gmail.com	Male	7
3720	Benjamin	Molina		36	Ben.Molina@gmail.com	Male	2
5621	Estefania	Ortega	Master's & above	57	Est.Ortega@gmail.com	Female	2
24147	Edgardo	Torres	Bachelor's	46	Edg.Torres@gmail.com	Male	2
64076	Patricio	Ramirez	Below Secondary	21	Pat.Ramirez@gmail.com	Male	2
21379	Anibal	Castro	Bachelor's	26	Ani.Castro@gmail.com	Male	4
23714	Juan	Ortega	Master's & above	40	Jua.Ortega@gmail.com	Male	7
64703	Edgardo	Torres	Bachelor's	28	Edg.Torres@gmail.com	Male	2
6596	Anibal	Gonzalez	Master's & above	38	Ani.Gonzalez@gmail.com	Male	2
7999	Pedro	Moreno	Bachelor's	31	Ped.Moreno@gmail.com	Male	4
5387	Edgardo	Molina	Below Secondary	23	Edg.Molina@gmail.com	Male	5
57989	Williams	Jimenez	Bachelor's	29	Wil.Jimenez@gmail.com	Male	2
26977	Luis	Hernandez		28	Lui.Hernandez@gmail.com	Male	2
63714	Alberto	Lopez	Master's & above	36	Alb.Lopez@gmail.com	Male	4
78556	Edgardo	Sanchez	Bachelor's	26	Edg.Sanchez@gmail.com	Male	2
51972	Juan	Perez	Bachelor's	32	Jua.Perez@gmail.com	Male	7
60572	Georgina	Moreno	Bachelor's	40	Geo.Moreno@gmail.com	Female	2
64	Juan	Molina	Bachelor's	42	Jua.Molina@gmail.com	Male	4
10732	Benjamin	Castro	Bachelor's	31	Ben.Castro@gmail.com	Male	2
9214	Anastacia	Diaz	Master's & above	35	Ana.Diaz@gmail.com	Female	3
10761	Isabel	Gonzalez	Master's & above	58	Isa.Gonzalez@gmail.com	Female	2
1555	Jorge	Ortega	Bachelor's	30	Jor.Ortega@gmail.com	Male	7
65450	Fernanda	Diaz	Master's & above	57	Fer.Diaz@gmail.com	Female	2

invoice_no	category	quantity	product_name	price	payment_method	invoice_date	store_id	customer_id
I299329	Toys	3	Plasticina	107.52	Cash	1970-01-01	10	C285588
I228197	Clothing	3	H&M T-Shirt	900.24	Credit Card	1970-01-01	7	C410550
I111278	Clothing	4	Adidas Ultraboost	1200.32	Credit Card	1970-01-01	1	C152494
I048394	Clothing	4	Under Armour HOVR	1200.32	Credit Card	1970-01-01	8	C265647
I257661	Clothing	3	Adidas Ultraboost	900.24	Cash	1970-01-01	8	C189333
I326825	Cosmetics	4	Clinique Moisturizer	162.64	Cash	1970-01-01	7	C193758
I862775	Cosmetics	5	Olay Regenerist	203.3	Credit Card	1970-01-01	3	C321450
I131103	Clothing	5	Levi's Jeans	1500.4	Cash	1970-01-01	5	C640986
I525641	Cosmetics	2	Olay Regenerist	81.32	Debit Card	1970-01-01	3	C738715
I172837	Toys	4	Rubik's Cube	143.36	Cash	1970-01-01	3	C145051
I152892	Clothing	5	Under Armour HOVR	1500.4	Credit Card	1970-01-01	7	C324250
I729395	Books	2	Moby Dick	30.3	Credit Card	1970-01-01	7	C481982
I738620	Souvenir	5	Shot Glass	58.65	Credit Card	1970-01-01	3	C320839
I190757	Clothing	1	Uniqlo Sweater	300.08	Cash	1970-01-01	6	C302132
I498695	Clothing	5	Puma Zapatillas	1500.4	Cash	1970-01-01	9	C445153
I322180	Clothing	3	Adidas Ultraboost	900.24	Credit Card	1970-01-01	8	C279286
I435103	Clothing	4	Zara Jacket	1200.32	Credit Card	1970-01-01	6	C227932
I472762	Clothing	1	Puma Zapatillas	300.08	Cash	1970-01-01	3	C902860
I161367	Cosmetics	4	Dove Soap	162.64	Cash	1970-01-01	4	C174660
I212712	Books	3	Jane Eyre	45.45	Cash	1970-01-01	1	C280767
I104000	Toys	5	Action Figure	179.2	Debit Card	1970-01-01	4	C824542
I264496	Clothing	2	Puma Zapatillas	600.16	Cash	1970-01-01	7	C356336
I949689	Cosmetics	1	Neutrogena Face Wash	40.66	Debit Card	1970-01-01	10	C238907
I214678	Clothing	4	Reebok Classic	1200.32	Cash	1970-01-01	2	C766797
I155069	Cosmetics	3	Garnier Micellar Water	121.98	Cash	1970-01-01	7	C143137
I339787	Cosmetics	4	Maybelline Mascara	162.64	Cash	1970-01-01	9	C259323
I170915	Toys	5	Board Game	179.2	Debit Card	1970-01-01	4	C264937
I205161	Clothing	2	Columbia Sportswear	600.16	Cash	1970-01-01	6	C195079
I099247	Clothing	5	Nike AIR Zapatillas	1500.4	Cash	1970-01-01	5	C166301
I125074	Books	3	Jane Eyre	45.45	Cash	1970-01-01	6	C197213
I162123	Clothing	2	Zara Jacket	600.16	Debit Card	1970-01-01	4	C782317
I331537	Clothing	2	Reebok Classic	600.16	Credit Card	1970-01-01	4	C243016
I274060	Clothing	5	Under Armour HOVR	1500.4	Debit Card	1970-01-01	4	C173368
I188854	Clothing	5	Under Armour HOVR	1500.4	Credit Card	1970-01-01	10	C217963
I610483	Clothing	1	H&M T-Shirt	300.08	Cash	1970-01-01	1	C169288
I183557	Clothing	4	Nike AIR Zapatillas	1200.32	Cash	1970-01-01	8	C304551
I140574	Souvenir	3	Fridge Magnet	35.19	Debit Card	1970-01-01	5	C200350
I200821	Toys	3	Action Figure	107.52	Cash	1970-01-01	6	C177910

Dashboard con datos de la base de datos:



<p>6. Intereses y proyecciones profesionales</p>	<p>El actual proyecto nos ayudó a profundizar en los conceptos de ciencias de datos, manejo de bases de datos, desarrollo web, estadísticas descriptivas, entre otros. Al desarrollar proyectos con mayor diversidad y abordando varios sectores y saber correlacionar los, nos proporciona mayor visión hacia donde nos queremos dedicar en el futuro, ya que estos trabajos son tan complejos que nos permiten tener que profundizar a grandes rasgos, todo lo que aprendimos durante la carrera.</p> <p>Al principio de la carrera, o a los primeros años, teníamos una visión diferentes sobre el mundo de la informática en general, pero al pasar de los semestres, encontramos varios sectores muy interesantes, como por ejemplo, el mundo de los datos, que es en donde nos basamos para construir nuestro proyecto, y en donde aguarda el corazón de nuestro sistema. Actualmente nuestras preferencias van más por este camino, de integración continua con múltiples plataformas de tecnología, interactuando con datos, desarrollo web, estadística, big data, etc.</p> <p>En el mundo laboral, planeamos empezar por estos sectores que manejamos, profundizando y buscando aprender cada vez más de este fascinante sector de los datos.</p>
--	---