**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 | *Admission Analytics* |
| Área (s) de desempeño(s) | * Big Data. * Ingeniería de Datos. * Ciencia de Datos. * Desarrollador * Ingeniería de Software |
| Competencias | * Desarrollar la transformación de grandes volúmenes de datos para la obtención de información y conocimientos de la organización, con el fin de apoyar la toma de decisiones y la mejora de los procesos de negocio. * Construir modelos de datos que soporten los requerimientos de la organización de manera escalable en el tiempo. * Gestionar proyectos informáticos, ofreciendo alternativas para la toma de decisiones de la organización. * Implementar soluciones sistemáticas integrales para automatizar y optimizar procesos de negocio de acuerdo a las necesidades de la organización. * Programar consultas y rutinas para manipular información de bases de datos, asegurando la disponibilidad de datos confiables y relevantes para la organización. * Ofrecer propuestas de solución informática analizando de forma integral los procesos de acuerdo a los requerimientos de la organización. * Elaborar proyectos innovadores que agreguen valor a contextos sociales y productivos, de acuerdo a las necesidades del entorno. |

|  |  |
| --- | --- |
| **Contenidos del informe final** | |
| 1. Relevancia del proyecto APT | Todo el proyecto pretende dar respuesta a una situación o problemática. En este caso, el proyecto APT busca solucionar la falta de información centralizada y detallada sobre los establecimientos educacionales, lo que limita la capacidad de las instituciones de educación superior en Chile para identificar y segmentar adecuadamente a los potenciales postulantes en un contexto altamente competitivo. Este desafío es relevante para el campo laboral de la ingeniería en informática, ya que implica el uso de Big Data, integración de sistemas, gestión de información y analítica para la toma de decisiones estratégicas. La situación se enmarca en el contexto nacional, donde universidades e institutos compiten por captar estudiantes de enseñanza media provenientes de diversos colegios, caracterizados por distintos niveles de matrícula y programas, y requieren información estratégica para diseñar campañas de admisión más efectivas. El problema impacta directamente a las instituciones educativas, que ven comprometidos sus ingresos al no alcanzar sus metas de matrícula; a las áreas de admisión, que necesitan información consolidada y oportuna; y a los propios estudiantes y sus familias, quienes se beneficiarían de una oferta educativa más clara y personalizada. El aporte del proyecto consiste en desarrollar una plataforma de Big Data que integre información externa de manera accesible, confiable y actualizada, permitiendo optimizar campañas de captación, mejorar la eficiencia de los procesos de admisión y fortalecer la competitividad institucional, contribuyendo además a que más estudiantes accedan a una educación alineada con sus intereses vocacionales. |
| 2. Objetivos | Desarrollar un sistema de analítica de datos que integre la limpieza, transformación y carga de información que consolide información externa sobre los establecimientos educacionales y sus cursos, con el fin de proporcionar al área de admisión una base de información accesible y actualizada que facilite la identificación y segmentación de potenciales postulantes, optimizando la toma de decisiones estratégicas para incrementar las matrículas y mejorar la competitividad institucional.   * Integrar y consolidar datos provenientes de fuentes externas, incluyendo información de establecimientos educacionales y sus cursos, en una plataforma centralizada de fácil acceso. * Diseñar un repositorio único de información que permita al área de admisión consultar datos relevantes de manera eficiente y en tiempo real. * Implementar herramientas de visualización y análisis que faciliten la segmentación de potenciales postulantes y el seguimiento del proceso de admisión. * Generar reportes e indicadores clave (KPIs) que apoyen la toma de decisiones estratégicas orientadas a mejorar las campañas de captación y el proceso de matrícula. * Fortalecer la competitividad institucional mediante el uso de datos confiables y oportunos para anticipar tendencias y orientar acciones en el mercado educativo. |
| 3. Metodología | Para abordar el problema identificado, se implementará una metodología basada en el ciclo de vida de un proyecto de Big Data, combinada con un enfoque ágil Scrum que permita entregar resultados parciales y usables al área de admisión en cortos ciclos de tiempo. Este enfoque asegura que la solución responda a las necesidades reales de la institución y se pueda ajustar de manera iterativa, garantizando flexibilidad y alineación con los objetivos del proyecto. Las etapas a considerar son las siguientes:   1. Levantamiento y análisis de requerimientos:  * Identificación de las necesidades específicas del área de admisión. * Definición de las fuentes de información internas (historial de matrículas, postulaciones, campañas previas) y externas (establecimientos educacionales, cursos, datos públicos del MINEDUC u organismos oficiales).  1. Recolección y consolidación de datos:  * Extracción de datos desde diversas fuentes internas y externas. * Normalización y limpieza de los datos para asegurar su calidad y consistencia. * Integración en un repositorio centralizado de información (data lake o data warehouse).  1. Diseño e implementación de la arquitectura Big Data:  * Selección de la infraestructura tecnológica adecuada (ejemplo: Hadoop, Spark, BigQuery u otra según disponibilidad). * Construcción de un repositorio único de información accesible para el área de admisión. * Establecimiento de procesos de carga y actualización automática de los datos (ETL/ELT).  1. Análisis y visualización de la información:  * Desarrollo de tableros de control (dashboards) y reportes interactivos. * Implementación de indicadores clave de gestión (KPIs) relacionados con postulaciones, captación y efectividad de campañas. * Segmentación de los potenciales postulantes según procedencia, colegio y cursos de interés. * La filtración de información necesaria  1. Validación y Documentación:  * Validación de la información y resultados con el área de admisión. * Entrega de documentación acerca de las características del sistema. * Validación de roles de acceso.  1. Evaluación y mejora continua:  * Monitoreo del impacto del sistema en la captación y conversión de estudiantes. * Identificación de mejoras y ajustes futuros para mantener actualizada y relevante la solución.   Respecto a la división de funciones y responsabilidades, el equipo de trabajo se ha organizado de la siguiente manera:   * Scrum Master: Esteban Rojas * Product Owner: Product owner * Scrum team: Antonio Vega * Dev Team: Esteban Rojas, David Rubio, Antonio Vega. |
| 4. Desarrollo | A lo largo del proyecto APT, desarrollamos como equipo las siguientes etapas y actividades principales:   1. Levantamiento de requerimientos y definición de alcance:    1. Revisión de la situación actual del área de admisión y sus necesidades de información.    2. Identificación de los indicadores clave de gestión (matrículas, postulaciones, procedencia de estudiantes, tipos de establecimientos, etc).    3. Priorización del alcance mínimo viable del proyecto (MVP), centrado en información de establecimientos y cursos. 2. Diseño del modelo de datos y arquitectura de la solución:    1. Elaboración de un modelo conceptual y lógico de datos para representar establecimientos, cursos, matriculas y variables asociadas.    2. Definición de la arquitectura general de la solución, incluyendo repositorio de datos, procesos ETL/ELT y capa de visualización.    3. Definición de roles de usuario(por ejemplo, administrador, manager y usuario) 3. Ingesta, limpieza y transformación de datos:    1. Carga de datos provenientes de fuentes externas (Datos abiertos MINEDUC)    2. Limpieza de datos: Tratamiento de valores nulos, duplicados, formatos inconsistentes y estandarización de identificadores.    3. Transformación de datos para adaptarlos al modelo definido (creación de tablas de hecho, de dimensiones, normalización) 4. Construcción del repositorio de datos y capa de consulta:    1. Implementación del repositorio único (data warehouse)    2. Creación de vistas y consultas para facilitar el acceso a la información por parte del área de admisión.    3. Optimización de consultas para mejorar tiempos de respuesta y experiencias de uso. 5. Desarrollo de la plataforma web Admision Analytics:    1. Implementación de la interfaz web para que los usuarios puedan acceder al sistema de forma amigable.    2. Configuración de autenticación y control de acceso por tipo de usuario.    3. Integración de los dashboards y visualizaciones de datos dentro de la plataforma. 6. Diseño de dashboards e indicadores (KPI):    1. Construcción de tableros con indicadores como: cantidad de establecimiento por comuna, tipo de establecimiento, matrícula por año, tendencia de matrícula, entre otros.    2. Incorporación de filtros por región, comuna, tipo de establecimiento y curso para apoyar la segmentación de potenciales postulantes.    3. Validación de los KPIs con la lógica del negocio de admisión. 7. Pruebas, validaciones y ajustes finales:    1. Ejecución de pruebas funcionales sobre la plataforma y las consultas de datos.    2. Corrección de errores detectados en la interfaz, la lógica de negocio o la calidad de datos.    3. Generación de la documentación final y preparación de la entrega del proyecto.   Durante el desarrollo del proyecto, hubo varios elementos que facilitaron el trabajo y permitieron avanzar de forma ordenada, tales como:   1. Experiencia previa en bases de datos, Google cloud y Power BI: El conocimiento previo en SQL, la plataforma de Google cloud, modelamiento de datos y herramientas de visualización ayudó a avanzar más rápido en el diseño del modelo y los dashboards. 2. Disponibilidad de datos estructurados: El acceso a fuentes de datos relativamente estructuradas facilitó la creación del modelo de datos y la construcción de indicadores. 3. Uso de enfoque ágil: La organización del trabajo en iteraciones, con tareas divididas por sprint, permitió ir entregando avances parciales, recoger feedback y realizar ajustes sin esperar el final del proyecto. 4. Trabajo colaborativo del equipo: La distribución de responsabilidades (Scrum Master, Product Owner y Dev Team) permitió que cada integrante se enfocara en tareas específicas, mejorando la coordinación y el avance.   También se presentaron diversas dificultades que fue necesario abordar:   1. Prácticas Profesionales y Trabajo: Dos de nuestros compañeros se encontraba realizando la práctica profesional, esto limitó los tiempos de trabajo. 2. Curva de aprendizaje de herramientas y tecnologías: La incorporación de nuevas tecnologías requirió un periodo de estudio y pruebas. 3. Cambios y ajustes en los requerimientos: A medida que se avanzaba, se identificaron nuevos indicadores y necesidades del área de admisión, lo que implicó revisar el alcance y priorizar que funcionalidades podían entrar en el MVP.   Para enfrentar las dificultades y asegurar el cumplimiento de los objetivos del proyecto, se realizaron los siguientes ajustes:  Priorización del alcance (MVP):  Se acotó el alcance inicial para concentrarse en un conjunto de indicadores críticos (por ejemplo, matrícula por establecimiento, comuna y tipo de establecimiento), dejando indicadores más avanzados para una fase posterior.  Refuerzo de la etapa de limpieza de datos:  Se incorporan rutinas adicionales de la validación y limpieza (revisión de duplicados, estandarización de nombres y códigos, controles de rangos válidos). Esto permitió mejorar la calidad de la información antes de cargarla al repositorio principal.  Reorganización de tareas en el equipo:  Se distribuyeron algunas responsabilidades para equilibrar la carga del trabajo (por ejemplo: separar claramente quien se encarga del modelo de datos, quien de la plataforma web y quien de los dashboards).  Simplificación de la interfaz y de los dashboards: A partir del feedback recibido, se simplifican algunos tableros para hacerlos más claros y directos, reduciendo la cantidad de gráficos por vista y destacando los KPIs más relevantes.  Documentación incremental:  En lugar de documentar todo al final, se optó por generar documentación a medida que se iban completando las funcionalidades, lo que ayudó a mantener claridad sobre las decisiones técnicas y de negocio. |
| 5. Evidencias | Las evidencias que se entregarán serán las siguientes:   * **Product Backlog:** Lista priorizada de funcionalidades y requerimientos del sistema, incluyendo detalles de cada historia de usuario. * **Historias de usuario:** Documento que detalla los escenarios de uso del sistema desde la perspectiva del usuario, con criterios de aceptación claros. * **Plan de Pruebas:** Registro de pruebas en el sistema, ya sea de funcionalidades o acciones realizadas en la plataforma web. * **Sprint Backlog:** Conjunto de tareas seleccionadas del Product Backlog que se abordarán en el sprint actual. * **Registro de Riesgos:** Registro de los riesgos identificados u ocurridos durante el proyecto. * **Arquitectura del sistema:** Diagrama y documentación detallada de la estructura del sistema, incluyendo módulos y conexiones. * **Burndown Chart:** Gráfica que muestra el trabajo pendiente en relación con el tiempo, permitiendo visualizar el progreso del proyecto y el ritmo del equipo. * **Kanban:** Herramienta visual que organiza el flujo de trabajo en columnas, mostrando tareas pendientes, en progreso y completadas, facilitando la gestión del avance del proyecto. * **Modelo de Datos:** Esquema del diseño en Data Warehouse y los tipos de datos que incluye. * **Impediment Log:** Registro donde se documentan los obstáculos que afectan el avance del proyecto, junto con su estado y acciones para resolverlos. * **Actas de Daily Meeting:** Registros breves de lo discutido en cada reunión diaria, incluyendo avances, impedimentos y próximos pasos del equipo. * **Sprint Review:** Reunión al final de cada sprint donde el equipo presenta el trabajo completado al Product Owner y stakeholders, evaluando el incremento y recopilando retroalimentación. * **Sprint Retrospective:** Reunión al final de cada sprint donde el equipo reflexiona sobre su desempeño, identificando lo que funcionó, lo que no, y proponiendo mejoras para futuros sprints. * **Costos:** Tabla de costos del proyecto, que incluye horas por persona y los recursos utilizados en el proyecto. |
| 6. Intereses y proyecciones profesionales | El proyecto APT fue una oportunidad concreta para profundizar nuestros intereses profesionales dentro del área de la Ingeniería en Informática, especialmente Big Data, ingeniería de Datos y analítica. En particular, el proyecto nos permitió:  Confirmar nuestros intereses por el trabajo con datos a gran escala: al diseñar el modelo de datos, implementar procesos de limpieza y construir indicadores, confirmamos que nos interesa trabajar en proyectos donde los datos son el eje central para la toma de decisiones.  Desarrollar habilidades en integración de sistemas y soluciones analíticas: la necesidad de integrar diferentes componentes (repositorio de datos, procesos ETL, plataforma web y dashboards) reforzó el interés por las soluciones completas de analítica, no solo por el desarrollo de plataforma tradicional.  Acercarnos a un contexto real de negocio: Al enfocarse en el área de admisión de instituciones educativas, el proyecto nos permitió entender mejor cómo la tecnología puede apoyar directamente objetivos estratégicos, como la captación de estudiantes y el uso eficiente de recursos.  Fortalecer el trabajo colaborativo y la metodología de proyectos: Trabajar con roles definidos nos ayudó a valorar la importancia de las metodologías ágiles y de la comunicación dentro de un equipo de desarrollo.  Luego de terminar el proyecto PT, nuestros intereses profesionales se mantienen en la misma línea que se planteó al comienzo de la asignatura, pero ahora están más claros y focalizados.  A partir de la experiencia obtenida en este proyecto, nuestras proyecciones laborales se orientan a seguir profundizando roles relacionados con datos y analítica, tales como:  Ingeniero de Datos o Desarrollador de Power BI: Nos gustaría trabajar en equipos encargados de diseñar modelos de datos, construir pipelines de información y desarrollar tableros de indicadores que den soporte a la gestión de la organización.  Especialista en soluciones de analítica: El enfoque del proyecto en el ámbito de la educación mostro el potencial de la analítica de datos en este sector, por lo que nos interesa de igual manera seguir explorando oportunidades en instituciones educativas, pero también en otras industrias donde la información sea clave para la toma de decisiones.  Continuar formándonos en tecnologías de Big Data y Cloud: A mediano plazo, nos proyectamos fortaleciendo nuestros conocimientos en plataformas en la nube, almacenamiento masivo de datos y herramientas de orquestación de procesos de datos, con el objetivo de asumir roles más avanzados en el futuro.  En resumen, el proyecto APT no solo permitió aplicar los conocimientos adquiridos durante la carrera, sino que también ayudo a definir mejor el camino profesional que quiero seguir, orientado al diseño, construcción y explotación de soluciones basadas en datos que generen valor para las organizaciones. |