

| **1. Resumen avance Proyecto APT** |
| --- |
| A continuación, encontrarás distintos campos que deberás completar con la información solicitada. |

| Resumen de avance proyecto APT | Hasta el momento, he realizado los siguientes avances en mi proyecto de Análisis Predictivo de Tendencias utilizando R y RStudio:   1. **Recopilación y limpieza de datos:** He trabajado en la obtención de datos históricos sobre el consumo de productos de Copec, como cupones electrónicos, TCT y TAE. Se ha llevado a cabo la limpieza y el preprocesamiento de los datos utilizando las bibliotecas **dplyr** y **lubridate** para asegurar la calidad y la integridad de los datos antes de proceder con los análisis. 2. **Análisis exploratorio de datos:** He realizado un análisis exploratorio de los datos para identificar patrones iniciales de comportamiento entre los diferentes segmentos de clientes. Utilizando **ggplot2**, he generado gráficos que han permitido identificar tendencias clave y verificar la calidad de las variables utilizadas en el análisis. 3. **Desarrollo del modelo predictivo:** Se ha comenzado a implementar un modelo de machine learning basado en técnicas supervisadas, como regresión logística y árboles de decisión, utilizando la biblioteca **caret**. Esto ha permitido categorizar a los clientes en segmentos como "nuevos", "perdidos", "decrecientes", "crecientes" y "reactivados". 4. **Evaluación preliminar del modelo:** Hasta el momento, el modelo ha sido evaluado usando métricas como R² y RMSE para medir su capacidad predictiva, obteniendo resultados prometedores. A medida que se avanza, se seguirán ajustando los hiperparámetros del modelo para mejorar su precisión.   **Objetivos específicos cumplidos:**   * Recopilación y preprocesamiento de datos históricos de consumo. * Exploración y análisis inicial para la segmentación de clientes. * Implementación inicial de un modelo predictivo.   **Ajustes a los objetivos:**  No se han realizado cambios significativos en los objetivos ni en la metodología hasta ahora. Sin embargo, se está considerando incorporar un análisis en tiempo real para la identificación más rápida de cambios en el comportamiento de los clientes, lo que podría influir en los próximos pasos del proyecto. |
| --- | --- |
| Objetivos | El objetivo de este análisis es crear un modelo predictivo que identifique y anticipe las tendencias de comportamiento de los clientes en el uso de productos de Skip, como cupones electrónicos, TCT y TAE. Se realizará un análisis de datos históricos de consumo para categorizar a los clientes en segmentos clave (nuevos, perdidos, decrecientes, crecientes y reactivados) y comprender las dinámicas que impulsan estos comportamientos. |
| Metodología | Utilizando la metodología CRISP-DM |
| Evidencias de avance | En este informe de avance, presentaré las siguientes evidencias para demostrar el desarrollo del proyecto de Análisis Predictivo del Comportamiento de los Usuarios de Skip:   1. **Informe de Análisis Exploratorio de Datos**: Este documento incluye visualizaciones clave y análisis estadísticos que muestran los patrones iniciales de comportamiento entre los distintos segmentos de clientes de Skip. Esto justifica el avance al haber identificado las principales características y tendencias que servirán como base para la creación del modelo predictivo. 2. **Código del Modelo Predictivo Inicial**: Se incluirá el código implementado en R, donde se detallan las técnicas de machine learning utilizadas (como regresión logística y árboles de decisión), así como las métricas de evaluación preliminares (R² y RMSE). Este código demuestra que el desarrollo técnico del modelo ha avanzado y que estamos en la fase de ajuste y mejora de precisión. 3. **Dashboard**: Este dashboard interactivo permite visualizar los resultados obtenidos en la segmentación inicial de los clientes y su evolución en el tiempo. Proporciona una herramienta útil para que los tomadores de decisiones de Skip puedan observar el comportamiento de los segmentos de clientes de manera intuitiva.   Estas evidencias reflejan el desarrollo del proyecto, ya que demuestran que se han realizado avances significativos en la preparación de datos, implementación del modelo predictivo y creación de herramientas de visualización para la toma de decisiones.  **Resguardo de la Calidad del Proyecto:**  Para garantizar la calidad del Proyecto APT, he seguido los principios y mejores prácticas propias de mi disciplina:   * **Correcta Aplicación de la Metodología CRISP-DM**: He utilizado esta metodología ampliamente aceptada en ciencia de datos para guiar todas las fases del proyecto, desde la comprensión de los datos hasta el desarrollo del modelo predictivo. * **Calidad en los Datos**: Para asegurar la integridad y precisión de los resultados, se ha realizado un preprocesamiento exhaustivo de los datos, que incluye la limpieza de valores atípicos, el manejo de datos faltantes y la conversión correcta de las variables categóricas. * **Uso de Herramientas y Recursos Especializados**: La implementación del modelo y el análisis se ha llevado a cabo utilizando herramientas avanzadas como R, con bibliotecas especializadas para machine learning, así como plataformas para la visualización. Esto asegura que los resultados sean confiables y basados en tecnología robusta. * **Validación Cruzada del Modelo**: Para evitar sobreajuste y asegurar la generalización del modelo, he aplicado técnicas de validación cruzada en la evaluación de su desempeño.   **Evidencias Adjuntas:**   * Informe de análisis exploratorio de datos. * Código del modelo predictivo. * Dashboard preliminar. |
| **2. Monitoreo del Plan de Trabajo** |
| Examina cuidadosamente tu plan de trabajo, enfocándote especialmente en la columna de estado de avance y ajustes. |

| Plan de Trabajo | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Competencia o unidades de competencias | Actividades | Recursos | Duración de la actividad | Responsable[[1]](#footnote-0) | Observaciones | Estado de avance | Ajustes |
| *Nombra las competencias o unidades de competencias que se relacionan con las diferentes actividades requeridas para el desarrollo del proyecto APT.* | *Nombra las actividades que se necesitan para desarrollar el proyecto APT.* | *Nombra los recursos necesarios para llevar a cabo las actividades definidas.* | *Señala la duración de cada actividad.* | *Escribe el nombre del integrante del equipo responsable de cada actividad.* | *Señala las dificultades o facilitadores que se podrían presentar durante la ejecución de cada una de las actividades propuestas.* | *Describe el estado de avance de cada actividad.*  *Tipos de estado:*  *En curso/ Con retraso/ No iniciado/ Completado/ Ajustada* | *Señala los ajustes o reformulaciones que has realizado.* |
| 1.4.1 | La **Identificación y Tratamiento de Valores Faltantes y Anomalías** consiste en detectar y manejar los datos incompletos o inusuales en el conjunto de datos. Esto incluye identificar valores faltantes, anomalías o errores en los datos y aplicar métodos adecuados para corregir o imputar estos problemas, asegurando así la integridad y calidad del análisis. | **Herramientas de Análisis de Datos**: Software como Python (pandas, numpy) o R para detectar y tratar datos faltantes y anomalías.  **Métodos de Imputación**: Técnicas para reemplazar valores faltantes, como imputación por la media, mediana, o métodos avanzados.  **Documentación de Procedimientos**: Guías sobre cómo tratar diferentes tipos de anomalías y valores faltantes.  **Equipo de Datos**: Personal con habilidades en análisis y limpieza de datos. | 23-09-2024 hasta 25-09-2024 | Lucas Carrasco | **Dificultades:**  **Identificación de Anomalías**: Dificultad para detectar valores atípicos o errores en grandes volúmenes de datos. Solución: Utilizar herramientas de visualización y técnicas estadísticas para identificar anomalías.  **Elección del Método de Imputación**: Decidir el mejor método para manejar los valores faltantes puede ser complejo. Solución: Evaluar diferentes métodos de imputación y seleccionar el más adecuado basado en el contexto y el impacto en los datos. | En Curso | Buscar mejores técnicas de tratamientos de datos faltantes para buscar los mejor valores para los modelos predictivos |
| 1.5.1 | La **Decisión sobre los Modelos a Utilizar** implica seleccionar los algoritmos y técnicas de modelado más adecuados para el análisis predictivo en base a las características del conjunto de datos y los objetivos del proyecto. Esta actividad requiere evaluar diferentes modelos, como Árboles de Decisión, Random Forest, Regresión Lineal, entre otros, y elegir aquellos que mejor se adapten a las necesidades del análisis y proporcionen los resultados más precisos. | **Herramientas de Modelado**: implementar y evaluar diferentes modelos de análisis predictivo.  **Conjunto de Datos**: Datos históricos y transformados que serán utilizados para entrenar y evaluar los modelos.  **Documentación y Literatura**: Guías y artículos sobre los modelos disponibles, sus características y sus aplicaciones.  **Equipo de Datos**: Personal con experiencia en modelado predictivo y selección de algoritmos. | 14-10-2024 hasta 16-10-2024 | Lucas Carrasco | **Dificultades:**  **Evaluación de Modelos**: Determinar qué modelo es el más adecuado puede ser complicado debido a la variedad de modelos y parámetros. Solución: Utilizar técnicas de validación cruzada y pruebas comparativas para evaluar el rendimiento de cada modelo.  **Requerimientos Computacionales**: Algunos modelos pueden ser intensivos en recursos y tiempo de procesamiento. Solución: Optimizar el código y considerar modelos menos complejos si los recursos son limitados.  **Interpretación de Resultados**: La interpretación de los resultados de diferentes modelos puede ser difícil. Solución: Utilizar herramientas de visualización y análisis para facilitar la comprensión de los resultados. | Atrasado | Aún no se a podido adecuar el modelo debido a las variedad de modelos y parámetros |

| **3. Ajustes a partir del monitoreo** |
| --- |
| Profundiza en las observaciones de tu plan de trabajo. Analiza las actividades planificadas y señala qué aspectos facilitaron u obstaculizaron la ejecución del plan. Plantea cómo abordaste y/o abordarás los obstáculos. Por último, señala los ajustes que realizaste al plan de trabajo a partir de este análisis. |

| Factores que han facilitado y/o dificultado el desarrollo de mi plan de trabajo: *Describe los factores que han facilitado y/o dificultado el desarrollo de tu Proyecto APT hasta ahora. En el caso de las dificultades debes describir qué acciones tomaste y/o tomarás para solucionarlas.*  Falta de conocimientos en algunas partes del desarrollo del proyecto por eso me he demorado un poco más de la cuenta y falta de recursos como set de datos para poder comenzar a hacer el informe de análisis exploratorio la mejor solución para la falta de conocimiento fue buscar otro tipo de herramientas más fáciles de utilizar que cumplan con los requisitos pedidos para el desarrollo del proyecto como también recursos de material de clases para esas herramientas que ayudarán al proyecto para su entrega final |
| --- |
|  |

| Actividades ajustadas o eliminadas: *Señalar los ajustes que realizaste a tu plan de trabajo o actividades que eliminaste y, justifica por qué lo hiciste.*  *En el caso de que tu plan de trabajo no haya requerido ni requiera ajustes, justifica esta decisión a partir de los facilitadores que te han permitido desarrollarlo como fue planeado.*  Hasta ahora ninguna actividad a sido ajustada o eliminada del cronograma de proyecto hasta su entrega completa |
| --- |

| Actividades que no has iniciado o están retrasadas: *En caso de que* ***no hayas iniciado actividades o estén retrasadas*** *de acuerdo a tu planificación, señala los motivos por los que no has podido cumplir dichos plazos y qué estrategias utilizarás para avanzar en dichas actividades y no afectar tu proyecto APT.*  Desde la actividad 1.5.1 la cual es Decisión sobre los Modelos a Utilizar hasta 1.8.4 la cual es Preparación del informe y cierre final pero se tendrá todo el proyecto antes de la entrega estimada |
| --- |

1. En caso de que el Proyecto APT sea grupal, en esta columna deben indicar el nombre de los responsables de cada tarea o actividad. Esto posteriormente permitirá diferenciar la evaluación por cada integrante. [↑](#footnote-ref-0)