Tecnicatura en Ciencia de Datos e IA

Práctica Profesionalizante I



*Integrantes*:

Couto Cristian

Juncos Sebastián

Padrón Belén

Pinedo Ramiro

Rigoni Bárbara

PRESENTACIÓN

Para: Grupo Newsan

De: Couto Cristian, Juncos Sebastián, Padrón Belén, Pinedo Ramiro, Rigoni Bárbara

Fecha de Inicio: 3 de octubre, 2024

Fecha de Finalización: 7 de noviembre, 2024

Asunto:

Estimados,

Somos un grupo de estudiantes del Centro Politécnico Superior Malvinas Argentinas. Actualmente nos encontramos cursando el primer año de la carrera de Técnico Superior en Ciencia de Datos e Inteligencia Artificial, en la materia “Práctica Profesionalizante I”, nos reunimos con el propósito de formalizar nuestra colaboración con la empresa Newsan.

Nos sentimos entusiasmados por la oportunidad de colaborar con su proceso de mejora continua y bienestar de sus colaboradores. Aportando al análisis ergonómico y la optimización de procesos productivos.

Durante nuestras prácticas, tenemos la expectativa de participar en el procesamiento, análisis y visualización de datos relevantes como: Análisis de puestos de trabajo, Mapa de riesgos y Matriz de rotaciones. Generando indicadores clave, que permitan optimizar las rotaciones del personal, minimizando riesgos ergonómicos y mejorando la seguridad y eficiencia en los puestos de trabajo.

Para este proyecto, daremos inicio el 3 de octubre, finalizando el 7 de noviembre. Y esperando poder contar con la disponibilidad de tiempo del equipo Newsan para la presentación de resultados.

**Contexto de Trabajo:**

Newsan tiene como objetivo asegurar un ambiente de trabajo seguro y saludable, priorizando la mejora continua de sus procesos productivos. La ergonomía se ha identificado como un área clave para alcanzar este objetivo, ya que el análisis ergonómico permite detectar oportunidades de mejora y diseñar estrategias que minimicen riesgos laborales. El acceso a datos relevantes como el análisis de puestos de trabajo, el mapa de riesgos y la matriz de rotaciones nos brindará información valiosa para colaborar en la mejora de las condiciones laborales.

**Problemática a Resolver:**

El problema que se aborda en este contexto es la identificación y prevención de riesgos ergonómicos en el lugar de trabajo. Para ello, la empresa realiza un análisis ergonómico detallado que le permite:

1. Identificar factores de riesgo en diferentes tareas, tales como movimientos repetitivos, postura forzada, manipulación de cargas, entre otros.
2. Implementar medidas correctivas y preventivas, como la rotación de tareas o el uso de herramientas adecuadas, que mejores la seguridad y el confort de los trabajadores.
3. Optimizar las condiciones laborales para mantener un equilibrio entre la salud de los empleados y la productividad.

Uno de los enfoques clave es la implementación de una matriz de rotaciones, que permite distribuir de manera equilibrada las tareas físicas y repetitivas entre los empleados, reduciendo el riesgo de lesiones y mejorando la eficiencia operativa.

**Objetivos Específicos:**

* Procesar y analizar datos para identificar indicadores clave en la planificación de rotaciones del personal.
* Identificar combinaciones de puestos de trabajo que minimicen el riesgo ergonómico.

La colaboración en este análisis es esencial para mejorar el entorno laboral.

PRODUCT BACKLOG

**Sprint 1**

1. **Confirmación de equipos**

* **Descripción**: Definir roles y responsabilidades para cada miembro del equipo.
* **Criterio de aceptación**: los roles y tareas están claramente asignados y documentados.
* **Duración estimada**: 4 horas

1. **Creación de grupos de WhatsApp**

* **Descripción**: formar grupos de comunicación en WhatsApp para el equipo.
* **Criterio de aceptación**: Todos los miembros del equipo están agregados y utilizan activamente el grupo.
* **Duración estimada:** 1 hora

1. **Creación de espacios en Google Drive**

* **Descripción**: organizar documentos en Google Drive para la colaboración del equipo.
* **Criterio de aceptación**: estructura de carpeta organizada, y todos tienen acceso.
* **Duración estimada**: 3 horas

1. **Reuniones por Google Meet**

* **Descripción**: programar reuniones periódicas por Google Meet para discutir avances y resolver dudas.
* **Criterio de aceptación:** calendario de reuniones compartido con todo el equipo.
* **Duración estimada**: 2 horas (programación + primera reunión)

1. **Creación de repositorios en GitHub**

* **Descripción**: crear repositorios para el control de versiones del código en GitHub.
* **Criterio de aceptación**: repositorios creados y todos los miembros con acceso adecuado.
* **Duración estimada:** 4 horas

1. **Creación del diagrama de Gantt**

* **Descripción**: elaborar un diagrama de Gantt para planificar y visualizar tareas y plazos.
* **Criterio de aceptación:** Diagrama de Gantt compartido y accesible para todos los miembros del equipo.
* **Duración estimada**: 6 horas

1. **Definir herramientas a utilizar**

* **Descripción**: seleccionar y documentar las herramientas tecnológicas que se usarán.
* **Criterio de aceptación:** lista de herramientas aprobada por todos los miembros.
* **Duración estimada:** 3 horas

1. **Documentación de presentación**

* **Descripción**: preparar la documentación formal del proyecto (objetivos, plan de trabajo, cronograma.
* **Criterio de aceptación**: documentación completada y revisada por el equipo.
* **Duración estimada:** 4 horas

1. **Elaboración del diccionario de datos**

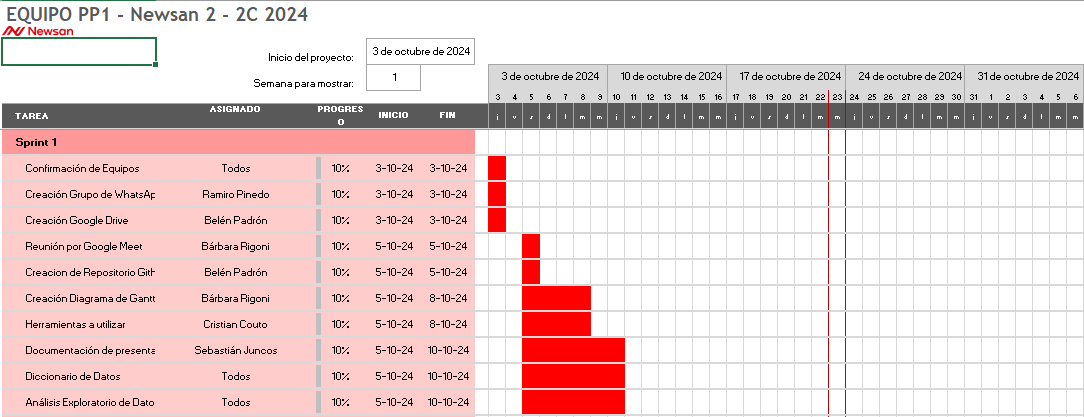
* **Descripción**: crear un diccionario que describa los elementos de datos del proyecto.
* **Criterio de aceptación:** diccionario de datos aprobado por el equipo.
* **Duración estimada:** 5 horas

1. **Análisis exploratorio de datos**

* **Descripción**: realizar análisis exploratorio para identificar patrones y características de los datos.
* **Criterio de aceptación**: informe del análisis completado y compartido con el equipo.
* **Duración estimada:** 6 horas

Trello/Diagrama de Gantt





PRODUCT BACKLOG

**Sprint 2**

1. **Limpieza y Normalización de Datos**

* **Descripción**: limpiar y normalizar los datos para eliminar valores atípicos y garantizar que sea uniformes y utilizables.
* **Criterio de aceptación:** los datos están libres de valores atípicos, duplicados o inconsistencias.
* **Duración estimada:** 5 días.

1. **Combinación de Datos**

* **Descripción**: Integrar y combinar diversas fuentes de datos para asegurar consistencia y coherencia.
* **Criterio de aceptación**: los datos de diferentes fuentes están correctamente combinados y se ha verificado su consistencia.
* **Duración estimada**: 5 días.

1. **Integración de APIs Externas**

* **Descripción**: conectar y consumir datos de APIs externas relevantes al proyecto.
* **Criterio de aceptación:** las APIs están integradas correctamente y los datos se están recuperando sin errores.
* **Duración estimada**: 6 días.

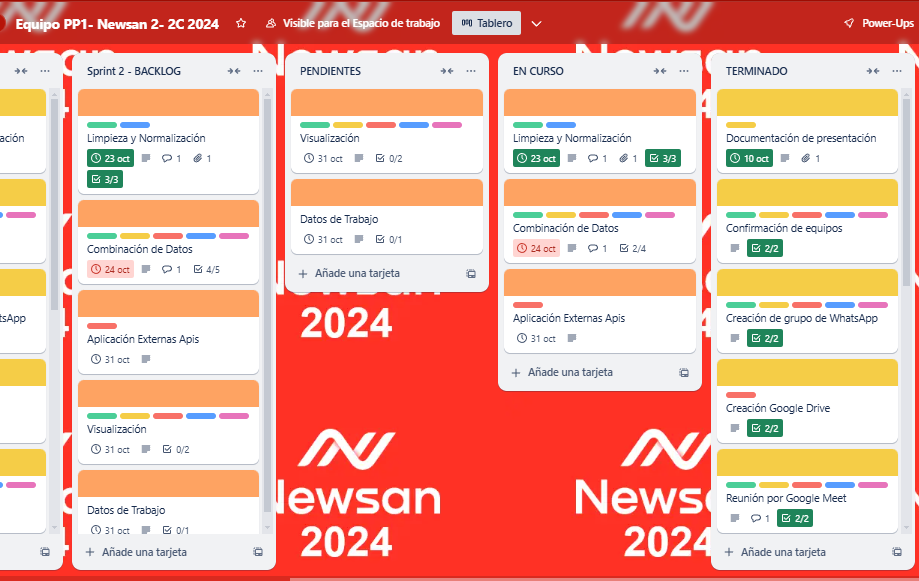
1. **Visualización de Datos**

* **Descripción**: crear visualizaciones iniciales para representar los datos combinados y procesados.
* **Criterio de aceptación**: visualización funcionales y claras, utilizando herramientas graficas (gráficos, tablas).
* **Duración estimada**: 5 días.

1. **Datos de Trabajo**

* **Descripción**: definir y estructurar los datos de trabajo finales que se utilizaran en el proyecto.
* **Criterio de aceptación:** los datos de trabajo están completamente definidos, estructurados y listos para ser utilizados.
* **Duración estimada**: 4 días.

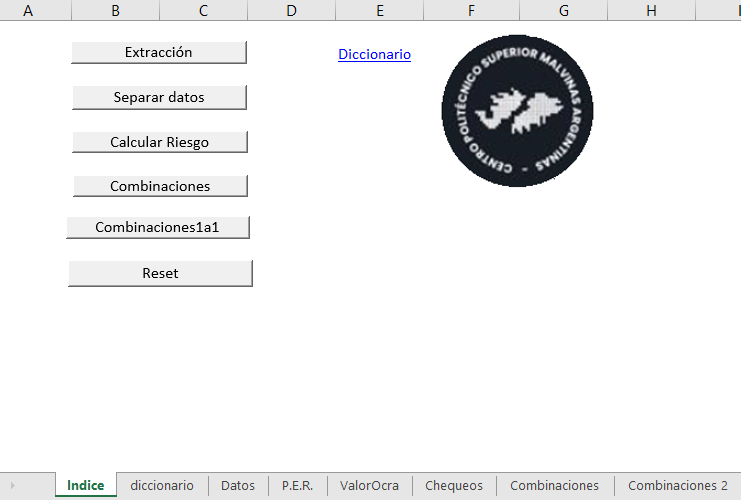
Trello/ Diagrama de Gantt

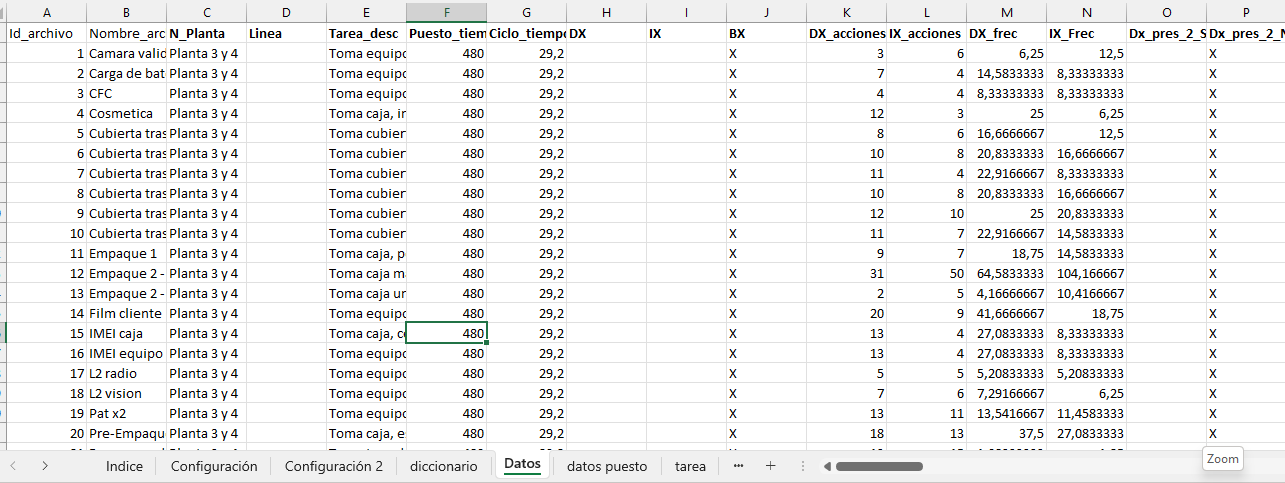




**ETL automatizada**

La disposición de los datos en el set presentaba la dificultad de que la extracción de los mismos solo se podía realizar manualmente, transcribiéndolos. Nos planteamos lograr que esa extracción pueda ser automática, lo que permitiría pasar de lograr un informe a una herramienta que podrá ser utilizada en el futuro con datos actualizados.



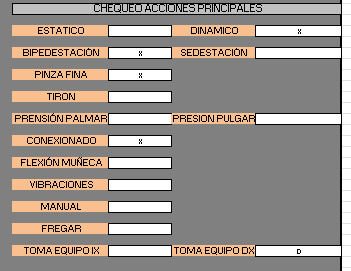


Como se puede observar en las capturas, esto se logró mediante la creación de 3 funciones-Macro cuyos parámetros se establecen en las pestañas “configuración” 1,2 y 3. Estas son:

* Extraer: la función busca en la carpeta seleccionada, dentro de todos los archivos .xlsx una pestaña “CHECK AUTOMATICO”, si no existe omite el archivo. SI la encuentra, crea una pestaña “datos” dentro del archivo fuente y allí vuelca los datos de las celdas que están indicadas en “Configuración” y le coloca los encabezados asignados allí también.
* Separar Datos: Esta función crea las pestañas indicadas en la pestaña “configuración 2” y copia desde la pestaña datos todos los registros, pero únicamente los campos cuyo encabezado se indica en la pestaña de configuración.
* Reset: Con esto se borran los datos de las pestañas indicadas en Configuración 3, con la excepción de los encabezados.

El objeto de este mecanismo/forma de lograr la extracción es lograr que en vez de crear un informe artesanal de los datos del set puedan actualizarse si hubiera cambios, o en el caso de agregar otros, o diferentes situaciones que pudieren suscitarse.

Asimismo, y en función de lo indicado por el Cliente, era necesario tipificar los tipos de movimientos que se ejecutaban en los diferentes puestos. Esta tipificación era en primer lugar, muy difícil de lograr, puesto que los datos incluían, en diferentes celdas, descripciones textuales de las diferentes partes de la tarea. Extraer esto implica depender de la formulación precisa y correcta de los textos, por lo que nos comunicamos con el cliente para acordar una tipificación de estos datos. Se nos proveyó de una lista de características de las acciones de cada puesto que era critico identificar. Nos pareció factible pensar en el ingreso directo de esta información mediante el formulario OCRA, que es el original del set de datos. Tanto los docentes como el cliente estuvieron de acuerdo. El resultado fue el siguiente:



El progreso del Sprint 2 siguió con el inicio de la generación de visualizaciones, modelado y combinación de datos.

Esta etapa fue demorada por cuanto carecíamos de la información necesaria para cumplir con los objetivos y requerimientos del cliente. Asimismo, fuimos citados a un encuentro con los docentes y el equipo Newsan 1, ya que del desarrollo de los sprints se

descubrió que no era necesaria la división del Set de datos original en dos, y se nos planteó unificar sets, diccionario y criterios.