

PRIMERA ENTREGA DEL PROYECTO

OBJETIVOS

- Proponer a una organización, una idea o problema a abordar desde la ciencia de datos
- Comprender la organización a partir de los datos disponibles para la propuesta

El propósito del proyecto del curso es aplicar la metodología de ciencia de datos y practicar lo aprendido en el curso en un entorno abierto. El grupo debe diseñar el proyecto en el que trabajarán y se realizarán tres entregas (tres sprints). En cada sprint se realizará una iteración de la metodología ASUM-DM, con énfasis en algunos pasos. Sin embargo, en todos los sprint, deben hacer una etapa de entender la organización y los datos, al igual que deben entregar un producto final.

DISEÑO DEL PROYECTO

El proyecto del curso debe cumplir con los siguientes requerimientos:

- Debe ser una oportunidad de aplicar ciencia de datos a una organización o para una población específica. La oportunidad debe estar soportada en términos de la organización, por ejemplo, de su definición estratégica (e.g., misión, visión, objetivos estratégicos, etc.), al igual que el entorno en el que la organización realiza sus negocios. También puede estar soportada en necesidades de la población objetivo y los objetivos de desarrollo sostenible.
- Debe poderse crear un modelo predictivo (Regresión o clasificación), es decir, se deben datos con una variable objetivo que se quiere predecir (o la forma de calcularla), al igual que ejemplos de registros con los posibles valores de esta variable objetivo.

DEFINICIÓN DE ROLES Y PRIMER SPRINT

Para el trabajo en equipo se sugiere asignar a cada integrante del grupo uno o más de los siguientes roles propuestos. Al final estos tres roles son los mínimos que se espera existan en los grupos de trabajo.

- **Líder de proyecto:** Está a cargo de la gestión del proyecto. Define las fechas de reuniones, pre-entregables del grupo y verifica las asignaciones de tareas para que la carga sea equitativa. Se encarga de subir la entrega del grupo. Si no hay consenso sobre algunas decisiones, tiene la última palabra.
- **Líder de negocio:** Es responsable de velar por resolver el problema o la oportunidad identificada y estar alineado con la estrategia del negocio o las características del grupo para el cual se plantea el proyecto. Debe garantizar que el producto se puede comunicar de forma apropiada al contexto del negocio o de la comunidad.

- **Líder de datos:** Se encarga de gestionar los datos que se van a usar en el proyecto y de las asignaciones de tareas sobre datos. Debe dejarlos disponibles para todo el grupo, y puede definir los datos que se usarán para cada iteración.
- **Líder de analítica:** Se encarga de gestionar las tareas de analítica del grupo. Se encarga de verificar que los entregables cumplen con los estándares de análisis.

Se recomiendan al menos las siguientes reuniones de grupo:

- *Reunión de lanzamiento y planeación del sprint:* Para definir roles y forma de trabajo del grupo. Se genera lluvia de ideas sobre el proyecto.
- *Reunión de ideación:* Una vez se han definido los datos que van a trabajar, la reunión de ideación busca definir el problema y los dolores.
- *Reuniones de seguimiento:* Se recomienda mínimo una reunión de seguimiento semanal corta. También pueden ser correos de avance según lo defina el grupo. Pueden tener un tablero de control de las tareas para lo cual se sugiere utilizar un software como Trello, Jira, Project.
- *Reunión de finalización:* Para consolidar el trabajo final, verificar el trabajo del grupo y analizar los puntos a mejorar para el siguiente sprint.

Se debe crear un repositorio para el proyecto del curso, en el que compartirán el código de la solución creada. Puede acceder a beneficios de estudiante en <https://education.github.com/pack>

ACTIVIDADES DEL SPRINT

En este sprint se realizará una iteración de la metodología ASUM-DM, con énfasis en los pasos de entender el negocio, enfoque analítico y entender los datos. Para lograr este objetivo, cada grupo debe seleccionar una organización, un problema o una población objetivo, ubicarla en un sector e incluir mínimo el siguiente contenido:

1. **[10%] Entender el problema:** Definir claramente la idea o problema a gestionar mediante la ciencia de datos. Entregar la información de negocio (estrategia, datos del sector, etc.) o de la población que sustente la relevancia de la idea o problema. Definir los objetivos del proyecto y métricas de negocio que se usarán para la validación.
2. **[10%] Ideación:** Incluir personas o arquetipos de usuarios, sus dolores y sus ideas. Se deben incluir las ideas de solución, la selección de una idea de producto y un prototipo de la solución esperada mediante un *story board*.
3. **[5%] Identificar implicaciones éticas, riesgos legales o sesgos que puede tener su proyecto.** Incluya consideraciones de privacidad y transparencia.
4. **[10%] Enfoque analítico:** Definir las hipótesis o preguntas a trabajar. Proponer los tipos de análisis predictivo que se realizarán a lo largo del proyecto.

5. **[15%] Entender los datos:** Describir las fuentes de datos a utilizar en función de su calidad e información que disponen para la gestión de la idea o problema.
6. **[40%] Entender los datos / producto:** Entregar un reporte del análisis descriptivo de las fuentes de datos, relevante para la validación de hipótesis. Validar las hipótesis iniciales.
7. **[10%]** Primeras conclusiones y acciones sugeridas para abordar la idea o solucionar el problema.

ENTREGA Y EVALUACIÓN

- El proyecto se realiza en grupos de 3 estudiantes.
- Documento por entregar
 - o Debe tener máximo 8 páginas (incluida tabla de contenido y página de presentación), a una columna y con letra arial tamaño 12.
 - o Deben especificar el enlace de github de su proyecto (puede ser un repositorio público o compartirlo a través de githubfront).
 - o Deben describir los participantes del proyecto y las contribuciones de cada uno.
- La fecha máxima de entrega es el **lunes 20 de septiembre a las 6:00 p.m.**