

## Proyecto Data Science

<b>Proyecto Data Science</b>	<b>1</b>
Indicaciones	2
Tema de investigación	2
Proyecto propio	3
Proyecto propuesto por Desafío Latam	3
<b>Hito 1 - Diseño</b>	<b>4</b>
Objetivo	4
Equipo de trabajo	4
Desarrollo de la investigación	4
<b>Hito 2 - Implementación</b>	<b>7</b>
Objetivo	7
<b>Hito 3 - Resultados</b>	<b>9</b>
Objetivo	9
<b>Hito 4 - Presentación Final</b>	<b>11</b>
Objetivo	11



**¡Comencemos!**

## Indicaciones

- Para realizar este proyecto debes haber cursado previamente todos los módulos de la carrera Data Science.
- El proyecto consta de 4 hitos (Diseño, Implementación, Resultados, Presentación Final), al terminar, comprime la carpeta de su proyecto con lo solicitado por cada hito y suba el `.zip` a la plataforma.
- El proyecto debe ser desarrollado en equipo, este debe estar compuesto por un mínimo de 4 y un máximo de 6 integrantes. **No se podrá trabajar de forma individual**, (a excepción de la grabación solicitada en el hito 4, que será individual).
- Puntaje Total del Módulo Proyecto Data Science: 40 puntos.
- Para aprobar el módulo, debes obtener al menos un total de 28 puntos.
- Existirá una calificación individual y una grupal en cada uno de los 4 hitos.

## Tema de investigación

Definir el tema de investigación del proyecto, este puede ser un proyecto propuesto por Desafío Latam o propio del equipo de trabajo. Se motiva a los equipos a trabajar en base a proyectos propios.

Para ambos tipos de proyectos, los resultados de la investigación pueden ser publicados o distribuidos en forma total o parcial por Desafío Latam para fines académicos u otros.

## Proyecto propio

- El proyecto debe ser real y puede estar asociado a algún requerimiento específico del trabajo de algún integrante del equipo o bien, enfocado en algún emprendimiento que se desee desarrollar.
- Se debe entregar la fuente de información, contacto que confirme dicha información, correo electrónico, teléfono de la empresa, entre otros antecedentes que permitan validar la fuente de información.
- Si la información es sensible, se debe realizar la encriptación de dicha información. Por ejemplo: rut, teléfonos, nombre de personas, entre otros.

## Proyecto propuesto por Desafío Latam

- Los proyectos propuestos por Desafío Latam deben ser con la data real y plantear la problemática que se desea resolver.
- El grupo de investigación podrá usar los datos y modificar o crear un nuevo alcance libremente.
- Desafío Latam no está obligado a entregar la fuente de información, sin embargo, para un mayor entendimiento debe explicar en forma genérica el origen y la problemática.
- La información entregada no debe tener datos sensibles, por ejemplo: rut, teléfonos, nombre de personas, entre otros.

## Hito 1 - Diseño

### Objetivo

Definir el equipo de trabajo, el tema de investigación, su motivación y plan de trabajo.

**Puntaje:** 10 puntos

### Equipo de trabajo

Creación del equipo de trabajo, definición de roles y responsabilidades de cada integrante. Un rol podrá ser asignado a más de un integrante.

Roles propuestos:

- Líder del equipo.
- Ingeniero de datos.
- Ingeniero de control de calidad y validación de datos.
- Ingeniero de análisis de datos.
- Ingeniero de modelamiento.
- Documentador.

*Se sugiere asignar funciones según fortalezas de cada integrante del equipo. La evaluación individual considera el rol definido para cada integrante del equipo.*

## Desarrollo de la investigación

El grupo debe tener previamente definido el/los requerimientos de su proyecto, así como los datos a utilizar. En base a la identificación de los requerimientos y los datos a disposición, el alumno debe ser capaz de proponer una estrategia analítica con los siguientes puntos que serán evaluados:

### 1. Organización y planificación del proyecto

- Creación del equipo y definición de roles y responsabilidades. **(1 Punto)**
- Planificación de la investigación, definir la pregunta de investigación, hipótesis de trabajo, definición del vector objetivo, estrategias analíticas a nivel descriptivo, modelación y predicción de trabajo si corresponde entre otros conceptos que sean relevantes para el desarrollo de la investigación, como por ejemplo: **(4 Puntos)**
  - **Encriptación de datos sensibles:** Qué datos serán transformados.
  - **Tratamiento de datos faltantes:** Qué técnicas se usarán o se eliminarán los registros, entre otros.
  - **Muestreo:** Se trabajará con la población o la muestra. Explicar el motivo de la decisión.
  - **Análisis descriptivo:** Medidas de dispersión, posición, distribución, gráficas, entre otros.
  - **Técnicas o modelos predictivos:** Series de tiempo, regresión lineal múltiple, regresión logística SVM, entre otros.

Los modelos candidatos a considerar deben contener una justificación y una agenda de entrenamiento declarada previamente.

En base a la estrategia analítica definida, considerar las métricas de evaluación del modelo. Las métricas deben justificarse.

- **Para problemas de clasificación:** Exactitud, precisión y recall.
- **Para problemas de regresión:** Al menor error cuadrático medio y error medio absoluto.

*Se recomienda redactar la investigación en términos simples y claros (no abusar del lenguaje técnico). Simpleza, precisión y conceptos claves.*

## 2. Documentación

Crear un documento en donde se detallan los pasos anteriores y solicitar aprobación del docente para trabajar en el proyecto escogido. (El documento debe tener máximo 6 hojas)

## 3. Exposición (5 Puntos)

Presentar la solicitud de la idea del proyecto con al menos la participación del 80% de los integrantes del grupo de trabajo presentes en la clase. La presentación grupal tendrá una evaluación de **3 Puntos** y la individual de **2 Puntos**.

Si el alumno no participa en la presentación su puntaje será de **0 Puntos**.

*Se dispondrá de 15 minutos para exponer el estado de avance del proyecto*

## 4. Entrega

Cada alumno deberá subir a la plataforma el documento que detalla la planificación del proyecto (documento) y la presentación, antes que comience la clase donde será presentado este hito.

## 5. Aprobación del proyecto

El docente deberá aprobar o sugerir incorporar, eliminar y/o modificar el proyecto para continuar con el siguiente hito al finalizar la exposición.

Si ninguno de los integrantes tiene el puntaje de 5 puntos en este hito, el docente se comunicará con el líder del proyecto quien deberá a la brevedad coordinar con el equipo de trabajo para realizar las correcciones que deben estar resueltas a más tardar el día que se presente el hito 2.

**Para problemas de regresión:** Al menos error cuadrático medio y error mediano absoluto.

## Hito 2 - Implementación

### Objetivo

Presentar el estado de avance del proyecto y las principales conclusiones o hallazgos.

**Puntaje:** 10 puntos

Se debe considerar en esta etapa al menos los siguientes aspectos:

#### 1. Presentación y resultados

- El volumen de datos debe ser tal, que sea posible implementar un proceso simple de trabajo de datos.
- En base a los requerimientos analíticos, la implementación del código debe permitir:
  - Ingesta de los datos al ambiente de trabajo acorde al requerimiento definido anteriormente.
  - Definición de atributos y depuración de valores perdidos, outlets, tokenización de texto y calibración de clases.
  - Entrenamiento/modelación de datos.
  - Creación de análisis descriptivo y gráficos.
- Todos los puntos respectivos a la implementación de código deben ser probados en los conjuntos de entrenamiento y validación externa.

#### **3 Puntos**

- Todos los puntos respectivos a la implementación deben estar en archivos auxiliares.
- Todos los modelos entrenados deben ser serializados (guardar modelo final)

#### **2 Puntos**

#### 2. Documentación

Si en el hito 1 se ha modificado alguna definición establecida, en este nuevo documento se deberá incluir los cambios y justificar esta acción.

*Máximo 10 páginas*

Si no se hace entrega de este documento, se entenderá que no existe ningún cambio a la planificación original y éste será la guía para la evaluación de este hito.

### 3. Exposición

Presentar la evolución del proyecto con la participación de al menos el 80% de los integrantes del equipo de trabajo.

La presentación grupal tendrá una evaluación de **3 Puntos** y la individual de **2 Puntos**.

Si el alumno no participa en la presentación su puntaje será de **0 Puntos**.

*Se dispondrá de 15 minutos para exponer el estado de avance del proyecto*

### 4. Entrega

Cada alumno deberá subir a la plataforma la presentación, códigos y el documento que contiene las definiciones (este último solo si fue modificado).



## Hito 3 - Resultados

### Objetivo

Presentar el estudio de estrategias y sugerencias ante una audiencia en clases

**Puntaje:** 10 puntos

Para este hito, los proyectos deben cumplir con los siguientes:

#### 1. Presentación Preliminar

Se debe considerar en esta etapa al menos los siguientes aspectos:

- Equipo de trabajo y roles.
- Presentar la problemática.
- Definir el "Modelo de Negocios".
- Solución analítica.
- Proponer mejoras al modelo o análisis.

La presentación sobre los principales resultados y recomendaciones en base a los requerimientos entregados tendrá una evaluación grupal de **3 Puntos** y una evaluación individual de **3 Puntos**. Si el alumno no participa en la presentación su puntaje será de **0 Puntos**.

En el caso de existir una comisión evaluadora, la nota grupal e individual se obtendrá mediante una media simple con los puntajes asignados por los integrantes del comité evaluador.

*Se dispondrá de 15 minutos para exponer.*

*Se dispondrá de 5 minutos para consultas del comité evaluador o público participante.*

## 2. Documentación

Se deben crear los documentos:

- **Resumen Ejecutivo:** Principales resultados de su trabajo en base a los requerimientos entregados. Este reporte debe tener un máximo de 13 páginas de extensión. Su redacción debe ser simple y evitar lenguaje técnico y/o sofisticado **(2 Puntos)**.
- **Documentación Técnica sobre cómo se generaron los resultados:** La extensión de este documento puede ser variada. Este documento debe permitir al lector poder realizar una réplica del trabajo realizado **(2 Puntos)**.

## 3. Entrega

Cada alumno deberá subir a la plataforma la presentación, resumen ejecutivo y documentación técnica antes de la sesión en la cuál se presentará el hito 3.

## Hito 4 - Presentación Final

### Objetivo

Presentar el producto final ante una audiencia y evaluación de un jurado

**Puntaje:** 10 puntos

Para este hito se pide presentar el proyecto frente a un jurado que lo evaluará. Los requerimientos y estructura de la presentación son los mismos que los mencionados para el hito 3 mencionado anteriormente.

La presentación sobre los principales resultados y recomendaciones en base a los requerimientos entregados tendrá una evaluación grupal de **8 Puntos** y una evaluación individual de **2 Puntos**. Quien no participe en la presentación obtendrá **0 Puntos**.

Los puntajes durante la presentación serán asignados de acuerdo a lo observado en cuanto a coherencia de la propuesta, dominio del tema y pertinencia de las respuestas a las preguntas hechas por la audiencia y el jurado

En el caso de existir una comisión evaluadora, la nota grupal e individual se obtendrá mediante una media simple con los puntajes asignados por los integrantes del comité evaluador.

*Se dispondrá de 15 minutos para exponer.*

*Se dispondrá de 5 minutos para consultas del comité evaluador o público participante.*

Adicionalmente, realiza y graba una presentación de entre 3 y 5 minutos, de forma individual respondiendo las siguientes preguntas (**1 Punto**):

- a. **¿Qué problemática fue detectada o te fue planteada para el desarrollo del proyecto?** Para responder, considera qué necesidades existen y quién las tiene.
- b. **¿Cómo tu proyecto satisface esa problemática o necesidad?** Para responder, describe tu aplicación y señala cómo lo que realizaste logra satisfacer la necesidad detectada.
- c. **¿Qué conocimientos y habilidades desarrolladas durante la carrera fueron claves para realizar este proyecto?** Para responder, identifica en los módulos anteriores aquellos aspectos técnicos y prácticos que aplicaste para el desarrollo de tu proyecto.
- d. Adicionalmente, reflexiona en torno a las siguientes preguntas: (ol style="list-style-type: none;">- i. **¿Qué dificultades tuviste para desarrollar la aplicación y cómo lo resolviste?** Para responder, recuerda aquellos tropiezos y

frustraciones, piensa qué estrategias o apoyos te permitieron salir adelante.

- i. **¿Qué fue lo que más disfrutaste de desarrollar tu proyecto?** Ya sea del proceso, del resultado o de aquello que te haya entregado mayor satisfacción.
- ii. **¿De qué manera crees que la metodología de aprendizaje fue un aporte para el resultado obtenido?** Para responder, mira hacia atrás y reflexiona sobre tu aprendizaje, la metodología de estudio, el trabajo colaborativo, entre otras cosas.



- Para realizar la grabación puedes usar cualquier aplicación de tu preferencia, te recomendamos [Loom](#), [Vidyard](#) o la herramienta de grabación de pantalla que tengas en tu PC.
- Sube el video a youtube o vimeo en modo público. Si prefieres, puedes marcarlo como no listado, de forma que solo quienes tengan el link puedan acceder a él.
- Puedes usar material de apoyo, incluso mostrar tu aplicación funcionando.
  - Carga un archivo con la URL del video en el LMS en el apartado '**Video proyecto**'.