

Sigla Asignatura	FMY0100	Nombre de la Asignatura	Fundamentos de Machine Learning	Tiempo	2h
Experiencia de Aprendizaje N° 1	Fundamentos de estadística				
Actividad N° 1.2	Recordando estadística				
Nombre del Recurso Didáctico	1.2.2 Actividad bibliotecas para graficar				

1. APRENDIZAJES E INDICADORES DE LOGRO

Aprendizajes (Procedimentales, Actitudinales y conceptuales)	Indicadores de logro
Analizar la problemática considerando los aspectos relevantes del proceso de negocio para establecer parámetros iniciales en la extracción y selección de datos.	Identifica las fuentes de datos asociadas a la problemática, con el objetivo de integrar y pre procesar la información.
Aplicar técnicas y herramientas en la preparación y visualización de los datos utilizando lenguaje Python u otro, de acuerdo a las buenas prácticas de la industria y necesidades del negocio.	Extrae los datos de las fuentes identificadas, de acuerdo a la problemática analizada.
	Realiza la limpieza, preparación e integración de ser necesario de los datos extraídos utilizando bibliotecas como Pandas, para obtención de la data preprocesada, de acuerdo al proceso KDD.
	Realiza la visualización de los datos utilizando bibliotecas como Matplotlib y Seaborn, para identificación de patrones de los datos.
Reconocer conceptos de estadística descriptiva orientada al manejo de datos, de acuerdo al tipo de herramienta utilizada.	Reconoce conceptos de estadística básica orientada al manejo de datos, de acuerdo al tipo de herramienta utilizada.
	Reconoce los diferentes modelos matemáticos aplicables al manejo de datos, según su uso en la industria.
	Diferencia algoritmos matemáticos aplicables al manejo de datos, según su uso en la industria.

2. DESCRIPCIÓN GENERAL ACTIVIDAD

1. Esta actividad tiene **carácter formativo**, es decir, tiene como propósito que experimentes lo revisado anteriormente en clases de forma práctica para, posteriormente, ser retroalimentado por tu docente e identificar aspectos de mejora.
2. Se dispondrá de material complementario relacionado con Python, donde conocerás las librerías más utilizadas en el análisis estadístico descriptivo y la creación de gráficos.

Paso 1

Crea un nuevo cuaderno para el desarrollo de tu laboratorio. Puedes usar un cuaderno de Jupyter de forma local o usar algún servicio en la nube. Te recomendamos usar la herramienta [Google Colab](https://colab.research.google.com/)¹.

Paso 2

A continuación, debes cargar el archivo de datos en un *dataframe* en tu cuaderno. El archivo que analizarás depende de tus preferencias, sin embargo, es importante que el análisis será desde un enfoque descriptivo y con gráficos (recomendación: tu archivo debe contar con variables cuantitativas).

Algunos sitios donde puedes encontrar datos para desarrollar la actividad:

- <https://www.kaggle.com/datasets>
- <https://data.fivethirtyeight.com/>
- <http://datos.gob.cl/en/>

Paso 3

Realiza un análisis de tus datos. Debes hacer una descripción de los datos global y por variable (al menos tres).

El análisis debe contar con los siguientes elementos como mínimo:

- Herramientas de resumen de análisis estadístico descriptivo (frecuencias, media, moda, percentiles, etc)
- Al menos dos tipos de gráficos en la representación de variables.

Paso 4

Finalmente, debes hacer al menos tres inferencias o deducciones basadas en tus datos. Debes justificar tus dichos de forma gráfica o descriptiva.

Aspectos a considerar en la evaluación

I. Utiliza un conjunto de datos apropiado y lo carga en su entorno de trabajo

- Escoge un conjunto de datos apropiado para el análisis (incluyendo variables cuantitativas)
- Carga los datos en el entorno usando los comandos y librerías adecuadas

II. Realiza un análisis inicial basado en estadística descriptiva y visualización

- Ejecuta los comandos adecuados para obtener un resumen estadístico de forma global y por variable
- Genera gráficos que permiten profundizar y aclarar el análisis de la información

III. Realiza conjeturas orientadas a la toma de decisiones basadas en el análisis de datos

¹ colab.research.google.com/

- Fundamenta sus conclusiones basadas en el análisis de datos
- Argumenta de forma gráfica o numérica el fundamento de sus conclusiones o conjeturas