

Módulo: B2 Análisis de Datos con Python 2020

Objetivo general
Al finalizar el módulo serás capaz de utilizar Python y sus librerías para realizar análisis más robustos aplicando modelos estadísticos y matemáticos que permitan encontrar patrones y relaciones en los datos con el fin de generar visualizaciones de análisis univariados, bivariados y multivariados con Seaborn y Matplotlib y aplicar modelos de regresión, clasificación y predicción.
Temario
<ul style="list-style-type: none"> ● Sesión 01: Estimados de Locación y Variabilidad ● Sesión 02: Introducción a la visualización de datos: Distribuciones ● Sesión 03: Exploración de Variables Categóricas y Análisis Multivariable ● Sesión 04: Correlaciones y Regresión Linear Simple ● Sesión 05: Distribuciones muestrales y técnicas de evaluación de modelos ● Sesión 06: Visualización de Datos Avanzada ● Sesión 07: Pruebas A/B y Procesamiento de Lenguaje Natural ● Sesión 08: Introducción a Machine Learning - Clasificación No Supervisada y Supervisada
Checkpoint
<p>Al finalizar este módulo serás capaz de aplicar los requerimientos a tu proyecto personal y cumplir con las actividades que se enlistan a continuación. Estos requerimientos y actividades se cumplirán dependiendo del alcance de tu proyecto; sin embargo, al final del módulo serás capaz de aplicar cualquiera de estos requerimientos y actividades para cualquier proyecto.</p> <p>Durante el Checkpoint, el experto te dará retroalimentación y orientación sobre cada entregable que hagas. Recuerda desarrollar y plantear de manera óptima tu proyecto para hacer que el proceso de retroalimentación y orientación sea más efectivo.</p> <p>Para esta sesión, mapea con la lista lo que estará incluido en tu proyecto y dáselo al experto para que te pueda guiar durante el desarrollo de este. Cualquier duda puedes consultarle al experto.</p> <p>Sesión 01: Estimados de Locación y Variabilidad</p> <ul style="list-style-type: none"> <input type="checkbox"/> Identificar los tipos de datos estructurados que existen. <input type="checkbox"/> Identificar valores típicos y atípicos. <input type="checkbox"/> Realizar cálculos estadísticos robustos. <input type="checkbox"/> Utilizar estimados de locación y variabilidad para describir las columnas numéricas de un dataset.

Sesión 02: Introducción a la visualización de datos: Distribuciones

- ☐ Identificar la distribución de un conjunto de datos junto con su importancia.
- ☐ Utilizar boxplots para visualizar distribuciones.
- ☐ Filtrar valores atípicos usando el Score de Rango Intercuartílico.
- ☐ Caracterizar nuestras distribuciones usando tablas de frecuencias, histogramas y gráficas de densidad.

Sesión 03: Exploración de Variables Categóricas y Análisis Multivariable

- ☐ Graficar un conjunto de datos agrupado de acuerdo a dos variables categóricas.
- ☐ Graficar un conjunto de datos numéricos agrupados de acuerdo a una variable categórica.
- ☐ Agregar anotaciones y títulos a las gráficas realizadas hasta el momento.
- ☐ Realizar gráficas de barras para explorar la distribución de variables categóricas.
- ☐ Realizar tablas de contingencia y gráficas con múltiples axes para explorar dos o más variables categóricas.
- ☐ Usar boxplots y violinplots para explorar variables numéricas segmentadas por variables categóricas.

Sesión 04: Correlaciones y Regresión Linear Simple

- ☐ Realizar análisis bivariado con variables numéricas.
- ☐ Identificar valores atípicos y decidir qué hacer con ellos.
- ☐ Explorar las relaciones existentes entre nuestras variables numéricas.
- ☐ Entrenar modelos de regresión lineal para realizar predicciones.

Sesión 05: Distribuciones muestrales y técnicas de evaluación de modelos

- ☐ Explorar las distribuciones muestrales de estadísticas de las variables numéricas en nuestro dataset.
- ☐ Practicar el entrenamiento de modelos de Regresión Lineal Múltiple.

Sesión 06: Visualización de Datos Avanzada

- ☐ Realizar nuevos tipos de gráficas que ayuden a explorar y entender mejor un conjunto de datos.
- ☐ Estilizar gráficas para que sean agradables a la vista y llamen la atención.

Sesión 7: Pruebas A/B y Procesamiento de Lenguaje Natural

- ☐ Utilizar patrones Regex para limpiar los datos estructurados.
- ☐ Utilizar el objeto FreqDist de nltk para hacer análisis estadístico del dataset.
- ☐ Realizar visualizaciones de los conteos de frecuencias utilizando gráficas de barras.
- ☐ Realizar visualizaciones de las distribuciones de frecuencias de las longitudes de las palabras o de las oraciones.

- ☐ Realizar nubes de palabras para detectar los temas más importantes del conjunto de datos.
- ☐ Realizar un análisis de sentimientos del conjunto de datos.

Sesión 8: Introducción a Machine Learning - Clasificación No Supervisada y Supervisada

- ☐ Aplicar los algoritmos de Clasificación No Supervisada y Supervisada.