|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sigla Asignatura | *MDY7101* | Nombre de la Asignatura | *Minería de datos* | Tiempo | *3 módulos* |
| Experiencia de Aprendizaje N° 2 | *Análisis de datos con Python* | | | | |
| Actividad N° 2.4 | *Evaluación Formativa* | | | | |
| Nombre del Recurso Didáctico | *2.4.1 Evaluación práctica* | | | | |

1. **Aprendizajes e indicadores de logro**

|  |  |
| --- | --- |
| Aprendizajes (Procedimentales, Actitudinales y conceptuales) | Indicadores de logro |
| Aplica metodologías para el desarrollo de proyectos de minería de datos, considerando las herramientas de estadística básica para aplicar en este tipo de proyecto. | Realiza la transformación de los datos, discriminando su relevancia en el análisis, de acuerdo a las características que posean los datos.  Realiza la exploración de los datos utilizando herramientas de estadística básica para determinar los valores relevantes. |
| Analizar los datos mediante la aplicación de modelos estadísticos, utilizando el lenguaje Python. | Realiza la transformación de los datos, utilizando el framework Anaconda, clasificándolos de acuerdo a sus características.  Aplica las métricas asociadas a estadística básica, para detectar los valores relevantes.  Grafica los resultados obtenidos a partir del trabajo con los datos, de acuerdo a las necesidades y características del cliente. |

1. **Descripción general actividad***.*
2. Esta actividad tiene **carácter formativa**, es decir: No lleva una evaluación con nota de por medio.
3. Es una actividad a realizar en grupo.
4. Dispone de los 3 módulos de la sesión para responder a la evaluación.
5. Deberá entregar un notebook en Jupyter Notebook con el código Python que responda a los requerimientos.
6. El nombre del archivo debe incluir su apellido paterno + apellido materno + nombre. Formato del nombre debe ser:

***EV\_EDA\_grupo\_nn.ipynb***

1. **Enunciado***.*

La data se encuentra en el enlace donde descargó esta evaluación. La descripción de la data

* Item\_Identifier: Identificador del producto.
* Item\_Weight: Peso del producto.
* Item\_Fat\_Content: Producto alto en grasas.
* Item\_Visibility: El % del área de exhibición total de todos los productos en una tienda asignados al producto en particular.
* Item\_Type: Familia del producto
* Item\_MRP: Precio máximo del retail o precio de lista
* Outlet\_Identifier: Clave única de la tienda
* Outlet\_Establishment\_year: El año en el que se estableció la tienda
* Outlet\_Size: Tamaño de la tienda
* Outlet\_Location\_Type: Localización de la tienda
* Outlet\_Type: Tipo de tienda
* Item\_Outlet\_Sales: Venta del producto en la tienda particular.

Considerando el set de datos que ha descargado se pide cumplir con los siguientes requerimientos:

1. Carga de datos
2. Análisis exploratorio inicial que incluya:
   1. Mostrar cantidad de observaciones y características (debe quedar la evidencia de que usted sabe de dónde sacar esa información)
   2. Medidas estadísticas básicas de las columnas, dependiendo de su tipo. Existen 2 tipos de variables (cuantitativas y cualitativas) y las medidas estadísticas son distintas para cada tipo; a eso se refiere este requerimiento. Deberá interpretar las medidas de una columna de cada tipo. La interpretación debe quedar en el contexto del negocio, no solo los valores.
   3. Limpieza y transformación de datos (Nulos, Imputación de Nulos, Transformación de variables categóricas)
   4. Proponer, al menos, 3 preguntas que puedan ser respondidas a través de agrupaciones. Esto implica el uso de la función groupby().
   5. Mostrar algunas tendencias, distribuciones de los datos. Deberá usar al menos 3 gráficos para mostrar las tendencias/distribuciones. Cada gráfico debe incluir un análisis de lo que muestra.

CONSIDERACIONES

* Cuando le piden realizar limpieza y transformación DEBE ser selectiva/o para limpiar y transformar las variables que tengan sentido para su análisis. Por ejemplo, si hay 20 variables categóricas, no debería transformar las 20; debe seleccionar, justificando su selección, y luego transformar.