

**Facultad de Ingeniería**

**Departamento de Ciencias de la Ingeniería**

**RETAIL ONLINE – PATRONES DE COMPRA**

Autor(es):

Sebastián Herrera L.

Cristian Meza C.

Diego Muñoz C.

Pedro Hurtado.

Marcelo Guzmán.

Profesor tutor:

Billy Peralta

Santiago, Chile

2020

**Definición de proyecto**

Este proyecto consiste en encontrar los patrones de compra dentro de un conjunto de datos de ventas online de Retail. Se implementarán estrategias para mejorar la velocidad del procesamiento de estos datos con la implementación de un pre-procesado del DataSet con el fin de minimizar las tuplas necesarias.

Como primera instancia, el presente informe mostrará una descripción general de los datos del DataSet. Luego, se mostrarán los resultados del análisis 1D y 2D sobre los datos, explicando su representació. Posterior a esto, se explicará el pre-procesamiento implementado con su fundamento, para finalmente mencionar los experimentos que plantearemos en este trabajo.

Tomar en consideración que la ejecución del programa fue realizada en sistema operativo Windows.

**Descripción de datos**

A continuación, se explicará cada columna del DataSet para un mejor entendimiento del proyecto:

Número de filas totales: 541909

Número de columnas totales: 8

invoiceNo: Número de transacción. Si comienza con una C, significa que es una transacción cancelada.

SotckCode: Código de identificación para cada producto.

Description: Nombre del producto.

Quantity: Cantidad del producto por transacción.

InvoiceDate: Día y hora en la que la transacción fue efectuada.

UnitePrice: Precio por unidad del producto.

CustomerID: Código de identificación del cliente.

Country: País donde reside el cliente.

**Análisis de datos**

Un análisis de los datos permite una mejor comprensión del contexto y situación en que se sitúa la entidad. Para esta sección, se tomó en consideración el conjunto de datos resultante luego del preprocesamiento.

**Análisis 1D**

El análisis 1D se manifiesta como una representación numérica o categórica del

comportamiento general de una variable, donde se puede utilizar métricas tales como media, mediana, moda, etc. según sea el caso.

A continuación, en la figura n°1, se muestra el resultado del análisis 1D sobre la Base de Datos pre-procesada previamente.