**Instituto Politécnico Nacional**

**Escuela Superior de Cómputo**

**Segundo Proyecto – Rstudio**

**Materia: Programación para Ciencia de Datos**

**Intregantes:**

* **De Luna Ocampo Yanina**
* **Perea Samaniego Jesús Giovanni**
* **// Ruíz Aguilar Alex Gerardo**

**Docente:**

* **Galindo Durán Cristal Karina**

**Grupo:**

* **3AM1**

**Fecha de entrega: 27/10/2021**

**ÍNDICE**

¿Por qué decidimos utilizar estos datos?…….,…….……………….... 3

Diagrama de flujo …………………………………….………………….. 3

Análisis estadístico ………………………………………………………. 6

**¿Por qué decidimos utilizar estos datos?**

Trabajar con una cantidad de datos grandes, no solo debemos hacer caso al tamaño de estos si no también a la variedad del contenido y a la velocidad con la que los datos se generan, almacenan y analizan. Se le conoce como las 3V, volumen, velocidad y variedad de los datos. Sin embargo, algunas organizaciones agregan una cuarta V, que es la veracidad de los datos.

Pensamos en estos datos porque consideramos interesante analizar cuántas personas pueden pasar por cada estación durante un día, un mes o un año. Estos datos útiles debide a que podemos obtener gráficas interesantes con el manejo de todos estos datos. Tienen las 4V que necesitamos o que podemos considerar para que nuestros datos puedan ser confiables.

Debemos recordar que estos datos se recogen de una Base de Datos y los patrones son expresiones que describen un subconjunto de datos.

Para esto debimos seguir una serie de pasos importantes para lograr obtener el análisis presentado. Los pasos que se deben seguir son:

1. Seleccionar el conjunto de datos o del subconjunto de variables o muestra de datos sobre los cuales realizaremos el descubrimiento.
2. La limpieza y el procesamiento de lo datos se compone de operaciones como la recolección de la información necesaria sobre la cual realizaremos el proceso, decidimos estrategias sobre la forma en la que manejaremos los campos de los datos no disponibles.
3. Seguimos con la reducción de los datos, en este paso utilizamos métodos de transformación para reducir el número de variables.
4. Elegimos los algoritmos y métodos para buscar patrones en los datos y así poder elegir los modelos más apropiados para su representación.

Es de suma importancia seguir estos pasos debido a que esto nos ayudará a mejorar nuestros análisis y obtener mejores resultados.

**Diagrama de flujo**

**Análisis estadístico**