****

**SERVICIO NACIONAL DE ADIESTRAMIENTO EN TRABAJO INDUSTRIAL**

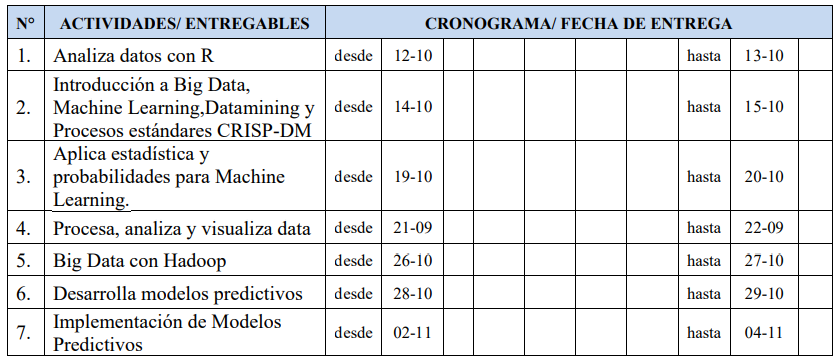
**PLAN DE TRABAJO**

**DEL estudiantE**

1. **INFORMACIÓN GENERAL**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Apellidos y Nombres: | Gamero de la Cruz Angeles | ID: | 1345941 | |
| Dirección Zonal/CFP: | Huancayo - Junin | | | |
| Carrera: | Ingeniería de Software con inteligencia artificial | Semestre: | | v |
| Curso/ Mód. Formativo | Big Data y Machine Learning | | | |
| Tema del Trabajo: | R studio | | | |

1. **PLANIFICACIÓN DEL TRABAJO**

****

# PREGUNTAS GUIA

**Durante la investigación de estudio, debes obtener las respuestas a las siguientes interrogantes:**

|  |  |
| --- | --- |
| **Nº** | **PREGUNTAS** |
| 1 | 1. ¿Cuales son las ventajas de utilizar Big Data y Machine Learning? |
| 2 | Qué aplicaciones existen en la industria de Big Data y Machine Learning? |
| 3 | ¿Cuales son la herramientas para Big Data y Machine Learning? |
| 4 | Cuales son las diferencias entre el modelo descriptivo e inferencial y el modelo predictivo e inteligente |

¿Que algoritmos existen el Machine Learning?

2

# HOJA DE RESPUESTAS A LAS PREGUNTAS GUÍA



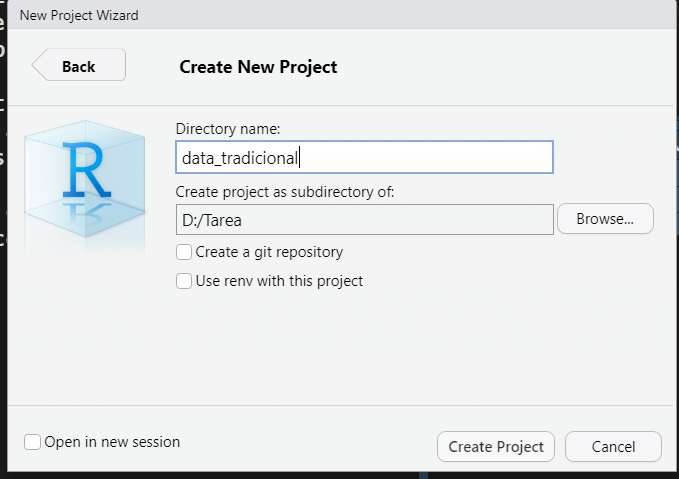
|  |  |
| --- | --- |
| 1. |  |
| Ventajas de usar Machine learning  --Mayor conocimiento de los clientes. Al contar con el Machine Learning, es posible determinar gustos, hábitos y necesidades de compra de los clientes. Esto mejora la experiencia del cliente y facilita su fidelización.  --Desarrollo del e-commerce. Mediante el conocimiento de los clientes, el Machine Learning puede determinar cuáles son los productos con mayor o menor demanda, así como las temporadas ideales para promociones y descuentos.  --Predicción de tendencias y necesidades. Permite anticipar los movimientos en la demanda, así como las necesidades que generará un producto o servicio, con la finalidad de desarrollar un auxiliar.  Ventajas de usa Big Data  --Velocidad en la toma de decisiones:  La información es fundamental como base para la correcta toma de decisiones, y mucho más cuando podemos manejar de forma dinámica toda la información que nos proporciona el Big Data. Podremos asumir decisiones inteligentes y veloces que ayuden a favorecer a nuestro negocio, puesto que éstas se han basado en un fundamento férreo. Es posible realizar de manera fluida un análisis de oportunidad antes de poner cualquier producto o servicio en el mercado.  --Planes estratégicos Inteligentes de Marketing:  A través de los datos que nos proporciona el Big Data, a día de hoy es posible analizar y predecir el comportamiento que un usuario tendrá en la red, conocer qué piensan los clientes sobre una marca o un producto, y cuáles son sus necesidades reales sobre la adquisición de productos o servicios. Se pueden analizar parámetros relativos al perfil específico de cada usuario, sus preferencias, sus tendencias o su vinculación a la marca, de forma que nos sea posible elaborar campañas dirigidas de marketing con un nivel alto de personalización. | |
| 2. |  |
| Aplicaciones machine Lear Ning  -Detección de rostro. Lo vemos en nuestros móviles.  -Anti-spam. Mediante el uso de tags.  -Anti-virus. Detectando software malicioso.  -Genética. En la clasificación de secuencias de ADN. -Forecast. De clima o tráfico.  Aplicaciónes Big Data  -Personalización y transparencia hacia el consumidor. ...  -Salud. ...  -Modelos de negocio más eficientes y automatizados. ...  -Sostenibilidad e igualdad. | |

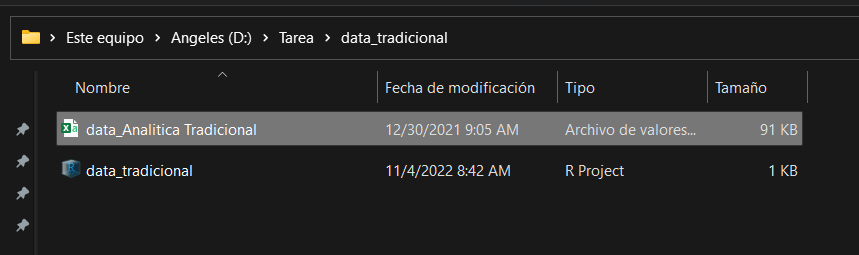
|  |  |
| --- | --- |
| 3. |  |
| -Herramientas de Big Data  -MongoDB. Dentro de las bases de datos NoSQL, probablemente una de las más famosas sea --MongoDB. ...  -Elasticsearch. ...  -Apache Spark. ...  -Apache Storm. ...  -Lenguaje R. ... -Python.  Herramientas de Machine learning  Scikit-Learn  Accord.net  TensorFlow  Weka | |
| 4. |  |
| Análisis Descriptivo: describe lo que ha pasado con estadísticas, gráficos, tablas e informes. Análisis Predictivo: realiza predicciones que van a ser útiles en el futuro. La calidad de la predicción es lo más importante. | |
| 5. |  |
| Algoritmos de regresión. ...  Algoritmos bayesianos. ...  Algoritmos de agrupación. ...  Algoritmos de árbol de decisión. ...  Algoritmos de redes neuronales. ...  Algoritmos de reducción de dimensión. ...  Algoritmos de Aprendizaje Profundo. | |

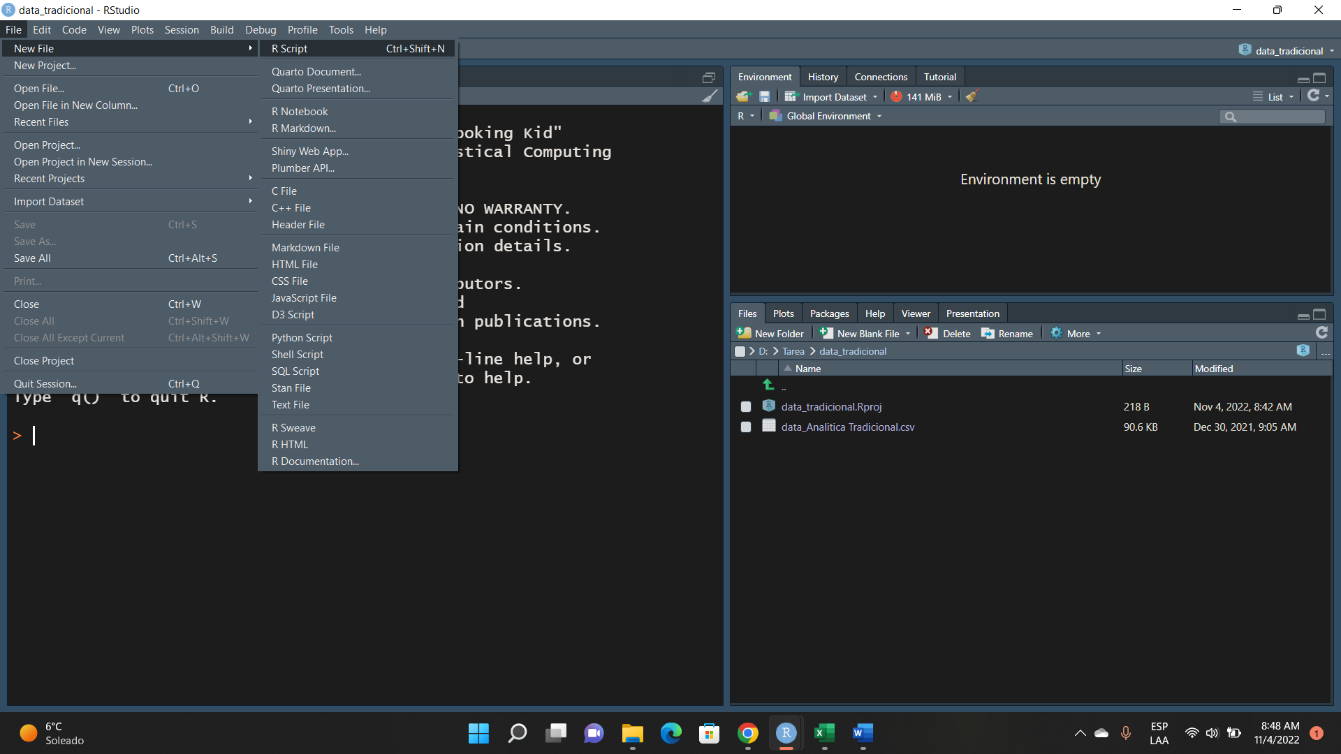
**PROCESO DE EJECUCIÓN**

|  |  |
| --- | --- |
| **OPERACIONES / PASOS /SUBPASOS** | **SEGURIDAD / MEDIO AMBIENTE / NORMAS -ESTANDARES** |
| **Instalar r estudio** | Norma Técnica ISO 27001:  Permite el aseguramiento y  confidencialidad de los datos  que se va a proporcionar el  sistema. |
| **Instalar package nnet y neuralnet (Data-1)** | Norma Técnica ISO 14000:  Esta norma garantiza que la  empresa va a obtener una  reducción de costos de energía,  implementando un uso  eficiente de los recursos  renovables. |
| **Instalar tidyverse, cluster, factoextra, nbclust (Data -2 .Data-3)** | Norma Técnica ISO 45001:  Mejorar nuestro rendimiento  laboral y trabajar más  activamente, también reducir  el tiempo de inactividad y el  coste de caída de servicio. |
| **Tabla de Excel en csv ( aplica en todos)** | Norma Técnica ISO 50001:  Permitirle una mejora continua  de la eficiencia energética, la  seguridad energética, la  utilización de energía y el  consumo energético. |
| **Instalar anaconda (Data – 4)**  **Ingresar a jupiter**  **Creamos una nuevo archivo en Python** | Norma Técnica ISO 9001:  Garantizar a sus clientes la  mejora de sus productos o  servicios y estos a su vez  prefieren empresas  comprometidas con la calidad. |
|  |  |

**INSTRUCCIONES:** debes ser lo más explícito posible. Los gráficos ayudan a transmitir mejor las ideas. No olvides los aspectos de calidad, medio ambiente y SHI.

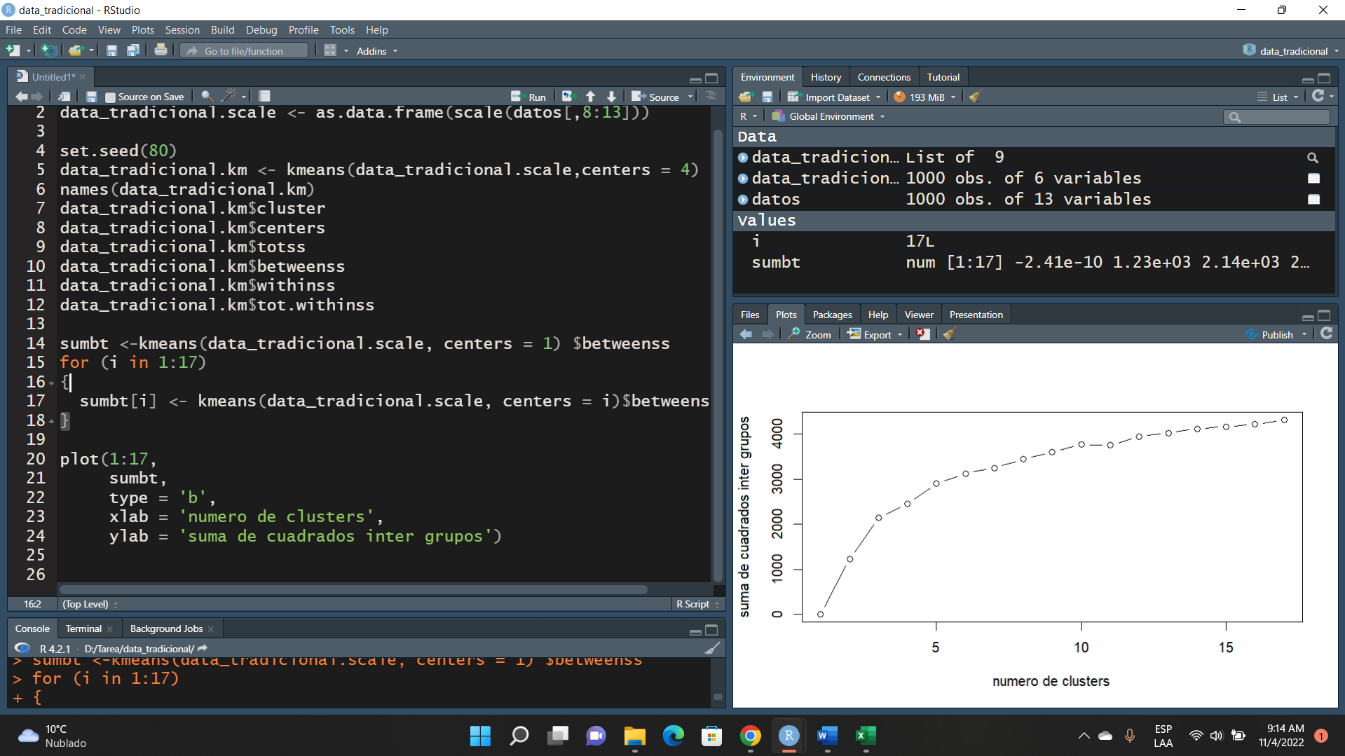
Paso 1:Después de instalar r estudio , abrimos y creamos un nuevo proyecto

Paso 2: pegamos el archivo ya brindado de Excel en rstudio

Paso 3:siguiente paso es generamos un script y lo guardamos

paso 4: Aquí vemos el número de clusters y la suma de cuadrados inter grupos

se calcula y grafica la suma de cuadros en cada numero de clusters, y allí se busca el cambio dependiente de empinada a poco profundidad, un codo para determinar el numero optimo de clústers



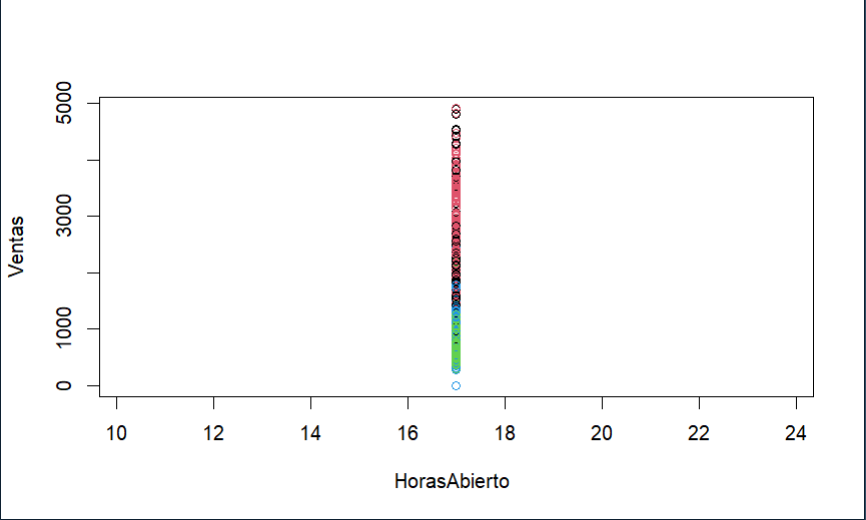
Paso 5:Análisis de los resultados

En este grafico podemos interpretar que mantienen las mismas cantidades de Horas Abiertas pero con diferencia de las ventas

En el año 2014 se evaluó que hubo la menor cantidad de ventas con una tasa aproximada de 621.3225 y también una creciente de ventas con la tasa aproximada de 2357.7224.

Interfaz de usuario gráfica

Descripción generada automáticamente



**DIBUJO / ESQUEMA/ DIAGRAMA**

|  |  |  |
| --- | --- | --- |
|  | | |
| C:\Unidad_D\Nuevo Logo\SENATI_FF-01 Modificado.jpg | [NOMBRE DEL TRABAJO] | |
| [APELLIDOS Y NOMBRES] | [ESCALA] |

**Ejecución de Modelo – Data 2**

Paso 1: Abrimos y creamos un nuevo proyecto

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Paso 2: pegamos el archivo ya brindado de Excel en rstudio

Captura de pantalla de un celular

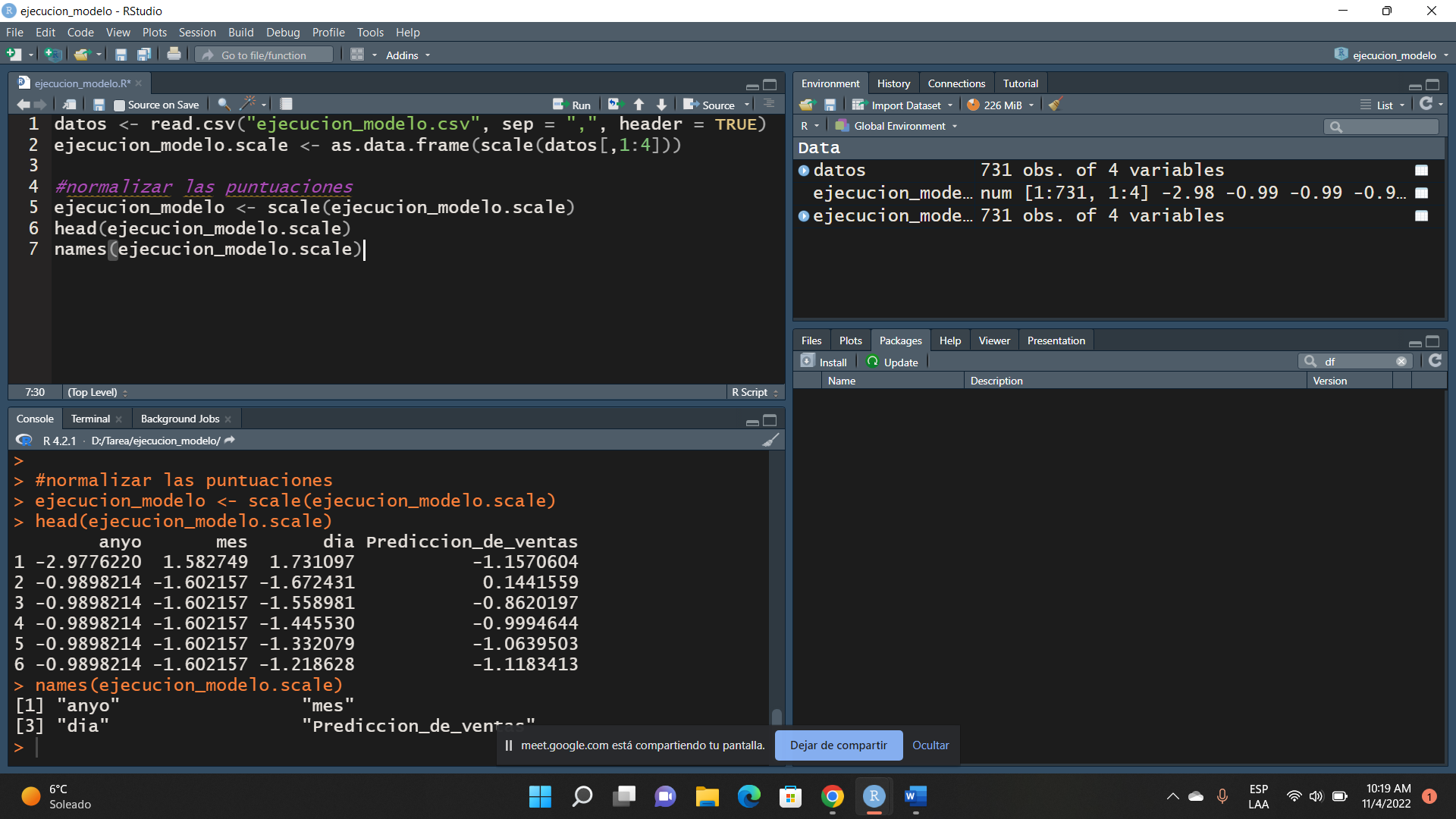
Descripción generada automáticamente

Paso 3:siguiente paso es generamos un script y lo guardamos

Captura de pantalla de computadora

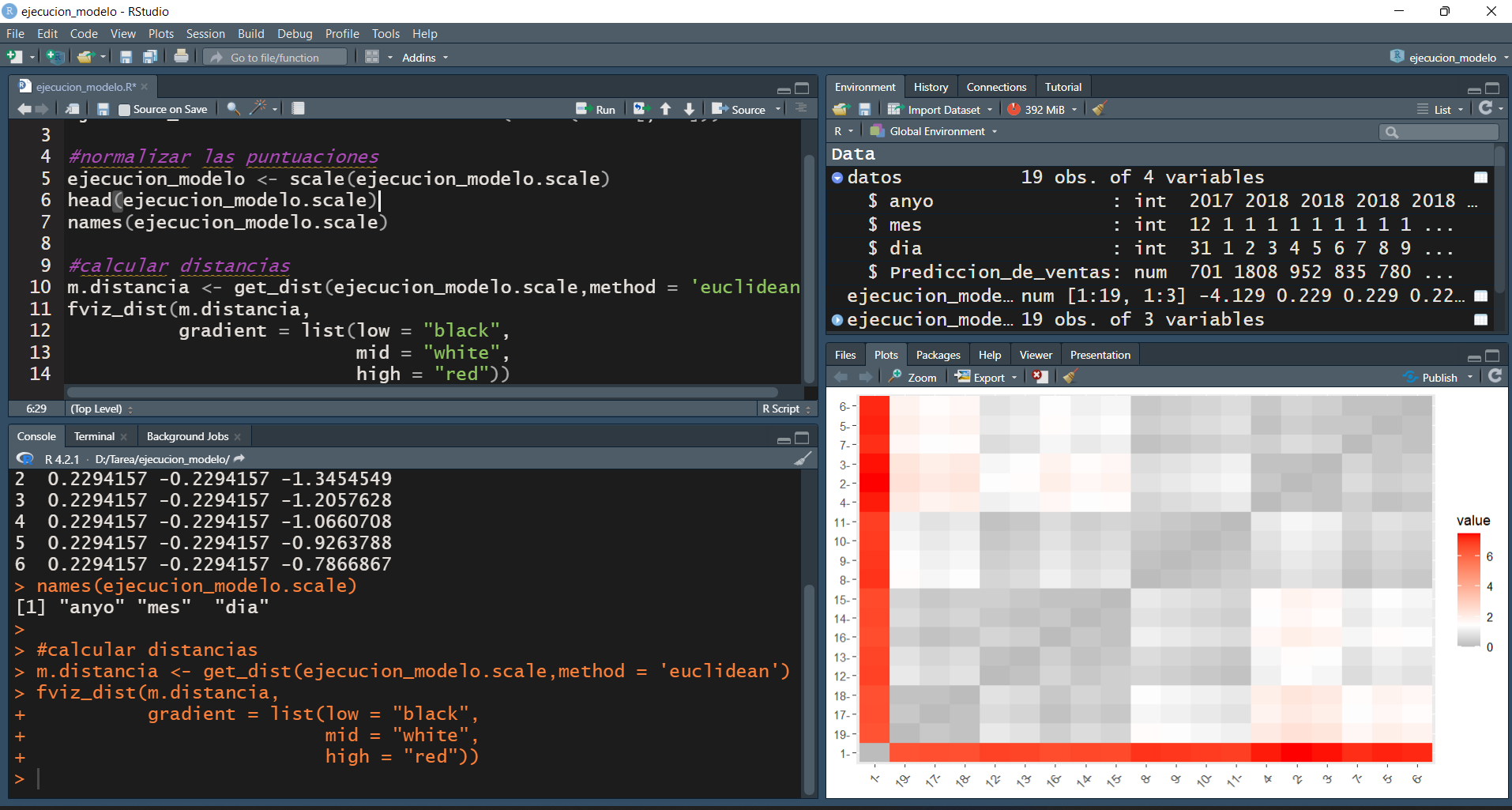
Descripción generada automáticamente

paso 4: Normalizar las puntuaciones



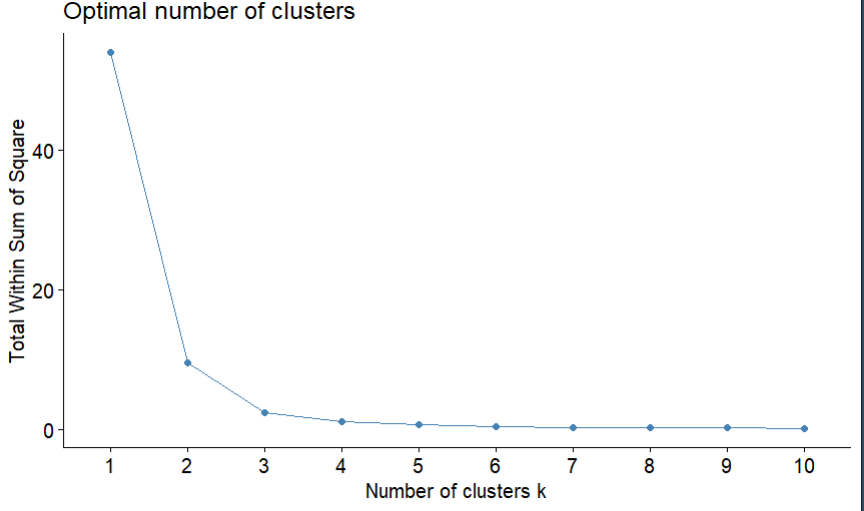
Paso 5:Análisis de los resultados calcular distancias

En este grafico podemos interpretar según los colores que observamos el rojo y blanco la distancia entre el año y el día según se ve en el valúes (grandes cantidades de datos)



Gráfico, Gráfico de líneas

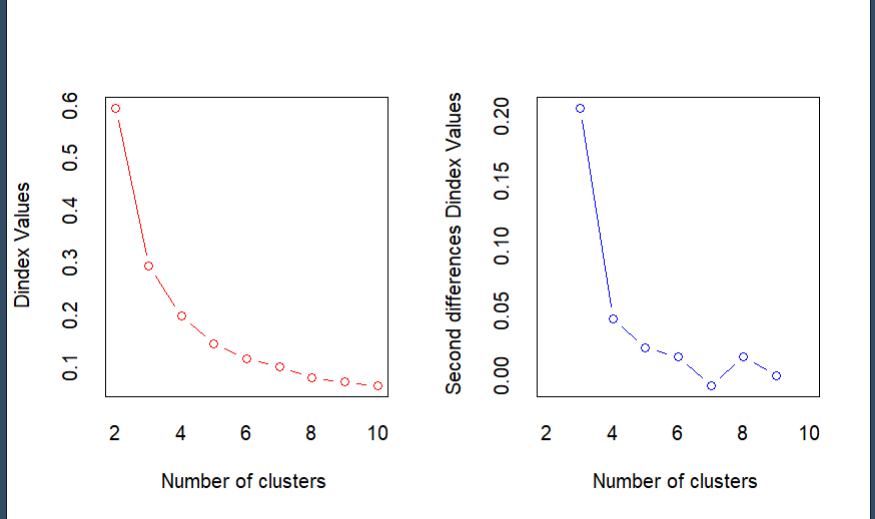
Descripción generada automáticamentePaso 6:estimar el numero de clusters



Gráfico, Gráfico de líneas

Descripción generada automáticamente

Paso 7: Comparamos el número de clusters con el dindex Valúes



**Evaluación Modelo – Data 3**

Paso 1:Abrimos y creamos un nuevo proyecto

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Paso 2: pegamos el archivo ya brindado de Excel en rstudio

Captura de pantalla de un celular

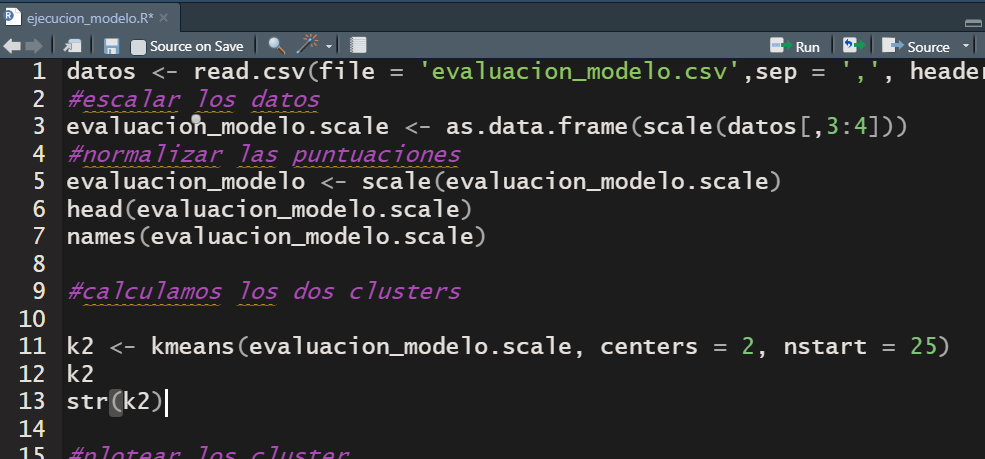
Descripción generada automáticamente

Paso 3:siguiente paso es generamos un script y lo guardamos

Captura de pantalla de computadora

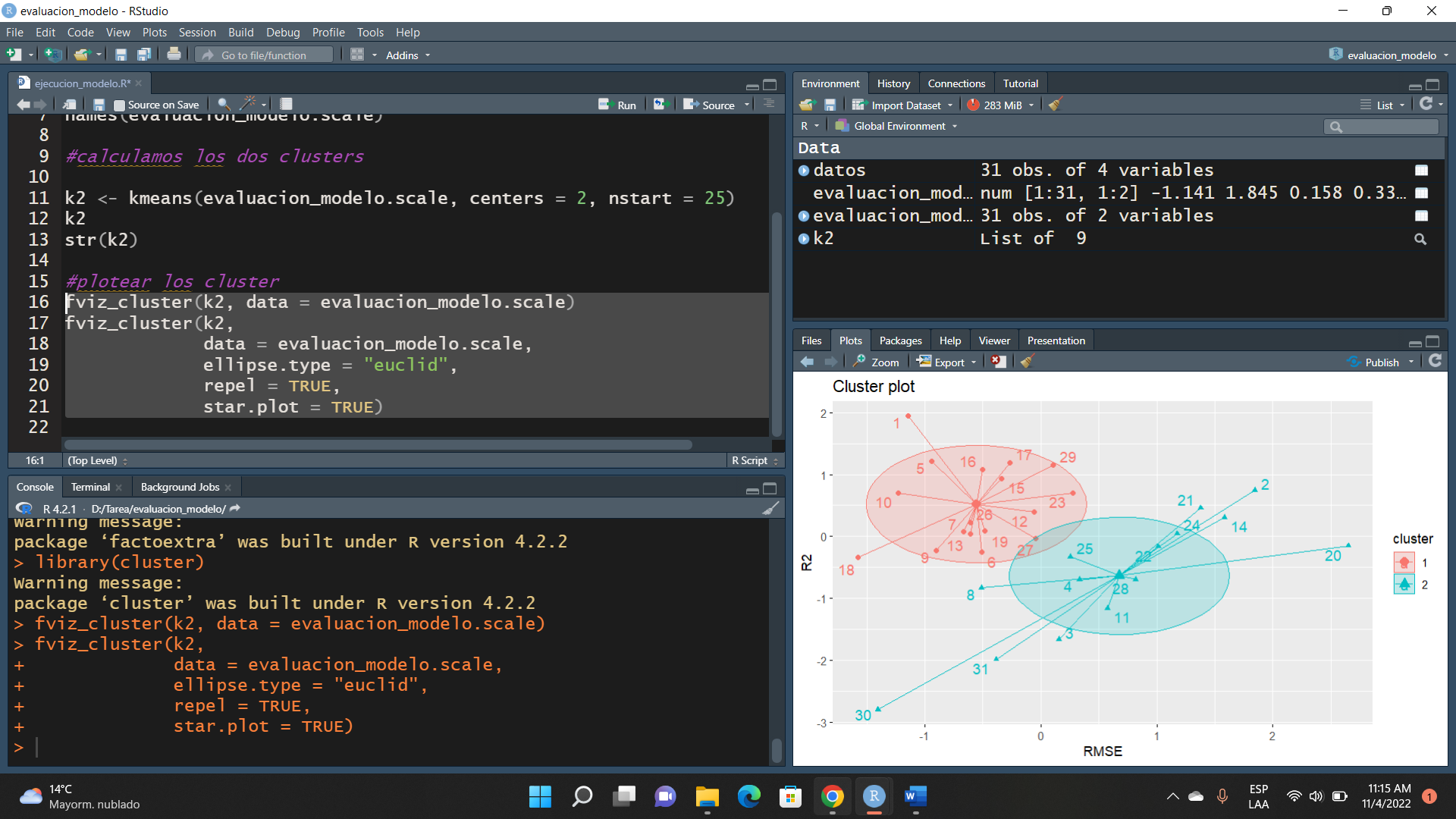
Descripción generada automáticamente

paso 4: Normalizar las puntuaciones



Paso 5:Los análisis de los resultados ,es el numero de clases a estimar de la distancia-medida-cuadrática-mínima , evaluando el coeficiente de determinación con el Conjunto de dependencias ubicadas en un campus .

Representado en las diferentes graficas



Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

Gráfico

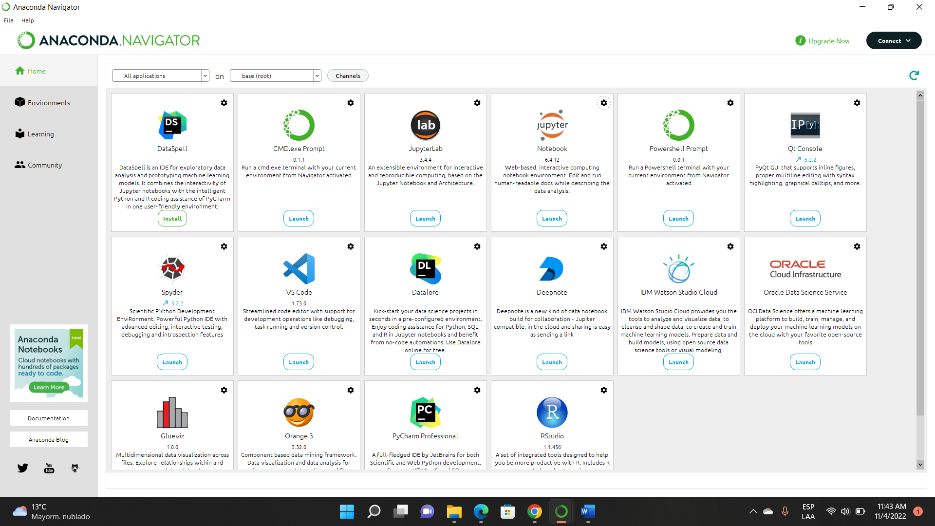
Descripción generada automáticamente

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

**Evaluación Modelo – Data 4**

Paso 1: Instalar Anaconda y ejecutar como administrador



Paso 2:Abrimos el launcher de Júpiter ,Creamos un nuevo archivo .py

Captura de pantalla de computadora

Descripción generada automáticamente

Paso 3:Ejecutamos el Código , Analizamos el grafico y podemos predecir la cantidad de ventas realizadas durante el año , mes y día como se muestra en la siguiente grafica.Imagen que contiene Calendario

Descripción generada automáticamente

**LISTA DE RECURSOS**

**INSTRUCCIONES: completa la lista de recursos necesarios para la ejecución del trabajo.**

|  |  |  |
| --- | --- | --- |
| **1. MÁQUINAS Y EQUIPOS** | | |
|  | laptop |  |
|  | mause |  |
|  | monitor |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **3. HERRAMIENTAS E INSTRUMENTOS** | | |
|  |  |  |
|  |  |  |
|  | rstudio |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **5. MATERIALES E INSUMOS** | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |