Universidad Tecnológica de la Costa

Ingeniería en Desarrollo y Gestión de Software

PORTAFOLIO DE EVIDENCIAS

Extracción de Conocimiento en Base de Datos

NOMBRE DEL PROFESOR:MARÍA EUGENIA GUERRERO CHAN

ALUMNO: PEÑA ORTIZ JOSE ALBERTO

24/05/20224

UT DE LA COSTA

Introducción

Contenido

Unidad 1: Introducción al análisis de datos

Resumen_U1_ECBD

Instrumento:	Resumen

Alumno: Peña Ortiz Jose Alberto	**Fecha:** 2 de mayo de 2024
Carrera: IDGS	**Grupo: IDGS91**
Asignatura: Extracción del Conocimiento en Base de Datos	**Unidad temática:** I. Introducción al análisis de datos
Profesor: MGTI María Eugenia Guerrero Chan	

Contenido

Instrucciones:

- Escribir una *Introducción* sobre las Metodologías para el análisis de datos (entre 200 y 250 palabras) y en *Desarrollo* la información que se está pidiendo que investigues.

Título: Metodologías para el análisis de datos

Contenido (Introducción y Desarrollo):

Introducción

En la era de la información, donde los datos se generan a un ritmo vertiginoso, el análisis de datos se ha convertido en una herramienta fundamental para extraer conocimientos valiosos y tomar decisiones informadas. Las metodologías para el análisis de datos proporcionan un marco estructurado y sistemático para abordar los desafíos que surgen al trabajar con grandes volúmenes de información.

Estas metodologías abarcan una amplia gama de técnicas y enfoques, desde el análisis descriptivo hasta el predictivo, pasando por la exploración de datos, el análisis inferencial y el análisis de texto, entre otros. Cada metodología tiene sus propias características y se adapta a diferentes tipos de datos y objetivos de análisis.

El análisis descriptivo se centra en resumir y describir los aspectos clave de un conjunto de datos, proporcionando una visión general de las tendencias y patrones presentes. Por otro lado, el análisis exploratorio de datos (EDA) permite profundizar en los datos, visualizarlos y descubrir relaciones y anomalías ocultas.

Cuando se trata de hacer inferencias y sacar conclusiones sobre una población más grande basándose en una muestra, el análisis inferencial entra en juego. Mediante técnicas estadísticas, como pruebas de hipótesis e intervalos de confianza, se pueden tomar decisiones informadas y evaluar la significancia de los resultados.

El análisis predictivo, por su parte, utiliza datos históricos y técnicas de modelado para hacer predicciones sobre eventos futuros. Esto es especialmente útil en áreas como el pronóstico de ventas, la detección de fraudes y la recomendación de productos.

Además, metodologías como el análisis de texto, el análisis de redes y el análisis espacial permiten abordar tipos de datos específicos y extraer información valiosa de fuentes no estructuradas o basadas en relaciones y ubicaciones geográficas.

Desarrollo

Que es un Proyecto de Análisis de Datos

Es el estudio profundo de una cantidad determinada de datos recopilados; se realiza con el fin de obtener conclusiones valiosas acerca de un aspecto en particular. Este estudio ayudará a la investigación del tema o aspecto y llevará a la toma de decisiones atinadas y a la creación de estrategias más eficientes

Características de un Proyecto de Análisis de Datos.

Un proyecto de análisis de datos se caracteriza por su enfoque multidisciplinario, su orientación hacia la resolución de problemas y su capacidad para trabajar con datos de diversas fuentes y formatos. Además, debe ser escalable y adaptable a medida que cambian las necesidades y los requisitos del proyecto.

Que es una Metodología para el análisis de datos.

Las metodologías de ciencia de datos proporcionan un marco sobre cómo proceder con los métodos, procesos y argumentos que se utilizarán para obtener respuestas o resultados y así tomar una buena decisión.

Una metodología para el análisis de datos es un conjunto de técnicas, herramientas y procedimientos utilizados para llevar a cabo el proceso de análisis de datos de manera sistemática y efectiva. Estas metodologías proporcionan un marco de trabajo estructurado que ayuda a organizar y ejecutar todas las etapas del proyecto de análisis de datos.

Principales Metodologías o Métodos de análisis de datos (definición de cada uno de ellos y ejemplos donde se aplican).

Análisis Descriptivo:

Definición: Implica resumir y describir las características clave de un conjunto de datos, como medidas de tendencia central (media, mediana, moda), dispersión (rango, varianza, desviación estándar) y distribución.

Ejemplos de aplicación: Informes de ventas, análisis demográfico, resúmenes estadísticos de datos de encuestas.

Análisis Exploratorio de Datos (EDA):

Definición: Se enfoca en explorar y visualizar los datos para descubrir patrones, tendencias, relaciones y anomalías. Utiliza gráficos, diagramas y técnicas de visualización.

Ejemplos de aplicación: Detección de valores atípicos, identificación de correlaciones, análisis de series temporales, segmentación de clientes.

Análisis Inferencial:

Definición: Implica hacer inferencias y sacar conclusiones sobre una población más grande basándose en una muestra de datos. Utiliza técnicas estadísticas como pruebas de hipótesis e intervalos de confianza.

Ejemplos de aplicación: Pruebas A/B, encuestas de opinión, estudios clínicos, investigación de mercado.

Análisis Predictivo:

Definición: Utiliza datos históricos y técnicas de modelado para hacer predicciones sobre eventos futuros. Emplea algoritmos de aprendizaje automático y modelos estadísticos.

Ejemplos de aplicación: Pronóstico de ventas, detección de fraudes, recomendación de productos, predicción de la rotación de clientes.

Análisis de Texto:

Definición: Implica extraer información y conocimientos a partir de datos de texto no estructurados. Utiliza técnicas de procesamiento del lenguaje natural (NLP) y minería de texto.

Ejemplos de aplicación: Análisis de sentimientos en redes sociales, clasificación de documentos, extracción de entidades, resumen automático.

Análisis de Redes:

Definición: Se centra en analizar las relaciones y conexiones entre entidades en una red o grafo. Utiliza medidas de centralidad, detección de comunidades y algoritmos de agrupamiento.

Ejemplos de aplicación: Análisis de redes sociales, detección de fraudes en transacciones, optimización de rutas, análisis de citas bibliográficas.

Análisis de Series Temporales:

Definición: Implica analizar datos que se registran a lo largo del tiempo para identificar patrones, tendencias y hacer predicciones. Utiliza técnicas como descomposición, suavizado y modelos ARIMA.

Ejemplos de aplicación: Previsión de demanda, análisis de tendencias de ventas, predicción de precios de acciones, pronóstico del clima.

Análisis Espacial:

Definición: Se enfoca en analizar datos geoespaciales para descubrir patrones y relaciones basados en la ubicación. Utiliza técnicas de sistemas de información geográfica (GIS) y análisis espacial.

Ejemplos de aplicación: Planificación urbana, análisis de crímenes, optimización de ubicaciones de tiendas, seguimiento de enfermedades.

Bibliografía

https://www.innovaciondigital360.com/periodista/equipo-editorial. (2023, July 18). Análisis de datos: Concepto, metodología y técnicas. InnovaciónDigital360; InnovaciónDigital360. https://www.innovaciondigital360.com/big-data/analisis-de-datos-tecnicas-y-metodologias-para-la-aplicacion-de-analytics/

ANÁLISIS DE DATOS. (2022). Tradeoff.mx. https://tradeoff.mx/2022/03/28/analisis-de-datos#:~:text=El%20an%C3%A1lisis%20de%20datos%20es,creaci%C3%B3n%20de%20estrategias%20m%C3%A1s%20eficientes.

Análisis de Datos | QuestionPro. (2023). Questionpro.com. https://www.questionpro.com/es/analisis-de-datos.html#:~:text=El%20an%C3%A1lisis%20de%20datos%20consiste,datos%20puede%20revelar%20ciertas%20dificultades.

Mildreth García. (2023, February 9). 5 metodologías de las ciencias de datos que te ayudará para tu estudio. Maestriasydiplomados.tec.mx; INSTITUTO TECNÓLOGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY. https://blog.maestriasydiplomados.tec.mx/5-metodologias-de-las-ciencias-de-datos-que-te-ayudar-para-su-estudio#:~:text=Las%20metodolog%C3%ADas%20de%20ciencia%20de,as%C3%AD%20tomar%20una%20buena%20decisi%C3%B3n.

ía ℂ. (2023, June 7). Gestionar un proyecto de datos en una empresa requiere un enfoque estratégico y una planificación cuidadosa. Implantar un proyecto de data en una compañía puede generar una amplia gama de beneficios significativos. Linkedin.com. https://es.linkedin.com/pulse/el-camino-hacia-%C3%A9xito-c%C3%B3mo-desarrollar-un-proyecto-de-garc%C3%ADa-cabria

Actividad 1: Ejercicios frecuencias

Crea la tabla de frecuencias, esta debe tener cada dato, sus frecuencias absolutas, frecuencias acumuladas, frecuencias relativas y frecuencias relativas acumuladas.

1. En una tienda de autos, se registra la cantidad de autos Toyota vendidos en cada día del mes de Setiembre.

0; 1; 2; 1; 2; 0; 3; 2; 4; 0; 4; 2; 1; 0; 3; 0; 0; 3; 4; 2; 0; 1; 1; 3; 0; 1; 2; 1; 2; 3

	Cantidad (A)	Frecuencia Absoluta (fi)	Frecuencia Acumulada (Fi)	Frecuencia Relativa (ni)	Frecuencia Relativa Acumulada (Ni)
0	0	8	8	0.266666667	0.266666667
0	1	7	15	0.233333333	0.5
0	2	7	22	0.233333333	0.733333333
0	3	5	27	0.166666667	0.9
0	4	3	30	0.1	1
0					
0					
0					
1					
1					
1					
1					
1					
1					
1					
2					
2					
2					
2					
2					
2					
2					
3					
3					
3					
3					
3					
4					
4					
4					
30	Total				
4	valor maximo				
0	valor minima				

2. Las calificaciones de los 23 estudiantes que tomaron la clase de matemáticas el año pasado son:

$$9,\ 10,\ 8,\ 10,\ 9,\ 8,\ 8,\ 9,\ 7,\ 9,\ 10,\ 8,\ 8,\ 7,\ 10,\ 8,\ 7,\ 7,\ 9,\ 9,\ 6,\ 9,\ 7.$$

	Calificaciones	Frecuencia Absoluta (fi)	Frecuencia Acumulada (Fi)	Frecuencia Relativa (ni)	Frecuencia Relativa Acumulada (Ni)
9	6	1	1	0.043478261	0.043478261
10	7	5	6	0.217391304	0.260869565
8	8	6	12	0.260869565	0.52173913
10	9	7	19	0.304347826	0.826086957
9	10	4	23	0.173913043	1
8					
8					
9					
7					
9					
10					
8					
8					
7					
10					
8					
7					
7					
9					
9					
6					
9					
7					
23	Total				
6	Valor Minimo				
10	Valor Maximo				

3. Un se está preparando para una maratón siguiendo una dieta muy estricta. A continuación, viene el peso en kilogramos que ha logrado bajar cada atleta gracias a la dieta y ejercicios.

0,2	8,4	14,3	6,5
4,6	9,1	4,3	3,5
6,4	15,2	16,1	19,8
12,1	9,6	8,7	12,1

	Peso Bajado	Frecuencia Absoluta (fi)	Frecuencia Acumulada (Fi)	Frecuencia Relativa (ni)	Frecuencia Relativa Acumulada (Ni)
0.2	0.2	1	1	0.0625	0.0625
8.4	3.5	1	2	0.0625	0.125
14.3	4.3	1	3	0.0625	0.1875
6.5	4.6	1	4	0.0625	0.25
4.6	6.4	1	5	0.0625	0.3125
9.1	6.5	1	6	0.0625	0.375
4.3	8.4	1	7	0.0625	0.4375
3.5	8.7	1	8	0.0625	0.5
6.4	9.1	1	9	0.0625	0.5625
15.2	9.6	1	10	0.0625	0.625
16.1	12.1	2	12	0.125	0.75
19.8	14.3	1	13	0.0625	0.8125
12.1	15.2	1	14	0.0625	0.875
9.6	16.1	1	15	0.0625	0.9375
8.7	19.8	1	16	0.0625	1
12.1					
16	Total				
0.2	Valor minimo				
19.8	Valor maximo				

Practica ejercicios_U1_ECBD

Instrumento	*Práctica de ejercicios*

Alumno: Peña Ortiz Jose Alberto	**Fecha:** 23 de mayo de 2024
Carrera: Ingeniería en Desarrollo y Gestión de Software	**Grupo:** IDGS91
Asignatura: Extracción de Conocimiento en Bases de Datos.	**Unidad temática:** I. Introducción al análisis de datos
Profesor: MGTI. María Eugenia Guerrero Chan	

I.- Ejercicios a resolver:

Instrucciones:

Observa la tabla siguiente e investiga la información que se pide y con base a la investigación llena la tabla.

VALOR (50%)

Tabla 1.- Llena la siguiente tabla con la información que se pide.

	Características	Casos de aplicación	Lenguajes y herramientas
Inteligencia Artificial			
Big Data			

Machine Learning			
Data Mining			

Con base a la información que pusiste en la tabla **selecciona** y **justifica** con tus propias palabras una herramienta de Inteligencia Artificial, Machine Learning, Data Mining y una de Big Data como apoyo para la toma de decisiones.

La justificación es de manera individual, es decir; una para Inteligencia Artificial, Machine Learning, Data Mining y una de Big Data (cantidad de palabras entre 250 y 300 o más si así lo requiere).

II.-Procedimientos y resultados: (Poner aquí la estructura y orden de la información)

	Características	Casos de aplicación	Lenguajes y herramientas
Inteligencia Artificial	1. **Automatización**: Capacidad de realizar tareas sin intervención humana. 2. **Adaptabilidad**: Capacidad de aprender y mejorar a partir de experiencias y datos. 3. **Reconocimiento de Patrones**: Identificación de patrones y correlaciones en grandes conjuntos de datos. 4. **Procesamiento del Lenguaje Natural (NLP)**: Entender y generar lenguaje humano. 5. **Visión por Computadora**: Análisis e interpretación de imágenes y videos. 6. **Toma de Decisiones**: Capacidad de tomar decisiones basadas en análisis de datos y modelos predictivos.	1. **Asistentes Virtuales**: Como Siri, Alexa y Google Assistant. 2. **Diagnóstico Médico**: Análisis de imágenes médicas y datos para detectar enfermedades. 3. **Automóviles Autónomos**: Vehículos que pueden conducirse sin intervención humana. 4. **Personalización en Comercio Electrónico**: Recomendaciones de productos basadas en el comportamiento del usuario. 5. **Detección de Fraude**: Identificación de transacciones sospechosas en tiempo real.	**Lenguajes** 1. **Python**: Popular por su simplicidad y gran cantidad de bibliotecas. 2. **R**: Utilizado principalmente en estadística y análisis de datos. 3. **Java**: Conocido por su robustez y capacidad de escalabilidad. 4. **Lisp**: Uno de los lenguajes más antiguos usados en IA. 5. **Prolog**: Utilizado en programación lógica y aplicaciones de IA. **Herramientas** 1. **TensorFlow**: Biblioteca de código abierto para el aprendizaje automático. 2. **Keras**: API de alto nivel para redes neuronales, que funciona sobre TensorFlow.

			3. **PyTorch**: Biblioteca de aprendizaje automático desarrollada por Facebook. 4. **OpenAI Gym**: Herramienta para desarrollar y comparar algoritmos de aprendizaje por refuerzo. 5. **IBM Watson**: Plataforma de IA que ofrece diversas herramientas y servicios.
Big Data	1. **Volumen**: Manejo de cantidades masivas de datos. 2. **Variedad**: Datos de múltiples fuentes y formatos. 3. **Velocidad**: Procesamiento rápido de datos en tiempo real. 4. **Veracidad**: Calidad y precisión de los datos. 5. **Valor**: Extracción de información útil y accionable.	1. **Análisis de Sentimiento**: Evaluación de opiniones en redes sociales y otros medios. 2. **Marketing Personalizado**: Campañas de marketing dirigidas basadas en análisis de datos. 3. **Análisis Predictivo**: Predicción de tendencias futuras y comportamiento del cliente. 4. **Optimización de la Cadena de Suministro**: Gestión eficiente de inventarios y logística.	**Lenguajes** 1. **Python**: Usado para análisis y manipulación de datos. 2. **R**: Preferido por estadísticos y científicos de datos. 3. **Java**: Fundamental en herramientas como Apache Hadoop. 4. **Scala**: Funciona bien con Apache Spark. 5. **SQL**: Lenguaje de consulta para bases de datos. **Herramientas**

		5. **Monitoreo y Mantenimiento Predictivo**: Prevención de fallos en equipos industriales.	1. **Apache Hadoop**: Marco para el procesamiento de grandes conjuntos de datos. 2. **Apache Spark**: Motor de análisis de datos rápido y de propósito general. 3. **HBase**: Base de datos NoSQL distribuida y orientada a columnas. 4. **Cassandra**: Sistema de gestión de bases de datos distribuido y escalable. 5. **Hive**: Herramienta de data warehousing construida sobre Hadoop.
Machine Learning	1. **Algoritmos Predictivos**: Modelos que anticipan resultados futuros. 2. **Aprendizaje Supervisado**: Entrenamiento de modelos con datos etiquetados. 3. **Aprendizaje No Supervisado**: Identificación de patrones sin datos etiquetados.	1. **Detección de Spam**: Identificación de correos electrónicos no deseados. 2. **Reconocimiento Facial**: Identificación de personas a partir de imágenes. 3. **Sistemas de Recomendación**: Sugerencias de productos, películas, etc.	**Lenguajes** 1. **Python**: Principalmente usado por su ecosistema de bibliotecas. 2. **R**: Extensivamente usado para análisis estadístico. 3. **Java**: Utilizado en producción por su robustez.

	4. **Aprendizaje por Refuerzo**: Modelos que mejoran mediante la experiencia y la retroalimentación. 5. **Generalización**: Capacidad de aplicar el conocimiento adquirido a nuevas situaciones.	4. **Análisis de Riesgo Crediticio**: Evaluación de la solvencia de los solicitantes de crédito. 5. **Control de Calidad en Manufactura**: Inspección automatizada de productos.	4. **Julia**: Conocido por su alto rendimiento en cálculos numéricos. 5. **MATLAB**: Usado en investigación y desarrollo académico. **Herramientas** 1. **scikit-learn**: Biblioteca de aprendizaje automático para Python. 2. **XGBoost**: Biblioteca para boosting de gradiente eficiente y flexible. 3. **LightGBM**: Biblioteca de boosting basada en árboles de decisión. 4. **CatBoost**: Biblioteca de boosting que maneja categóricas automáticamente. 5. **MLlib**: Biblioteca de aprendizaje automático para Apache Spark.
Data Mining	1. **Extracción de Conocimiento**: Descubrimiento de patrones y relaciones en grandes conjuntos de datos.	1. **Segmentación de Clientes**: Agrupación de clientes en base a características comunes.	**Lenguajes** 1. **Python**: Popular por sus bibliotecas como pandas y scikit-learn.

	2. **Análisis Descriptivo**: Resumen de las características de los datos. 3. **Análisis Predictivo**: Predicción de eventos futuros basados en datos históricos. 4. **Agrupamiento (Clustering)**: Clasificación de datos en grupos homogéneos. 5. **Reglas de Asociación**: Identificación de relaciones frecuentes entre variables.	2. **Análisis de la Cesta de la Compra**: Identificación de productos que se compran juntos. 3. **Detección de Anomalías**: Identificación de comportamientos anormales o fraudulentos. 4. **Optimización de Campañas de Marketing**: Mejora de la efectividad de las campañas basadas en análisis de datos. 5. **Análisis de Tendencias**: Detección de cambios en los comportamientos y preferencias del mercado.	2. **R**: Extensivamente usado en análisis de datos. 3. **SQL**: Fundamental para la extracción y manipulación de datos. 4. **Java**: Usado en muchas herramientas de minería de datos. 5. **SAS**: Utilizado en análisis estadístico y minería de datos. **Herramientas** 1. **RapidMiner**: Plataforma para el análisis avanzado y minería de datos. 2. **WEKA**: Conjunto de herramientas de aprendizaje automático para minería de datos. 3. **KNIME**: Plataforma de análisis de datos que permite crear flujos de trabajo. 4. **Orange**: Herramienta de minería de datos y visualización.

			5. **DataRobot**: Plataforma de aprendizaje automático automatizado.

Inteligencia Artificial: IBM Watson

IBM Watson es una herramienta poderosa para aplicaciones de Inteligencia Artificial. Desarrollada por IBM, Watson está diseñada para procesar y analizar datos no estructurados, lo que incluye texto, imágenes, audio y video. Una de sus características es su capacidad de procesamiento del lenguaje natural (NLP), lo que permite entender y responder preguntas en un lenguaje común, similar al humano. Esto es especialmente útil en aplicaciones como chatbots, asistentes virtuales y análisis de texto, donde la interacción fluida con el usuario es crucial.

Watson también ofrece una variedad de servicios de IA, como el análisis de sentimientos, el reconocimiento de imágenes y el procesamiento del habla. Estos servicios permiten a las empresas implementar soluciones avanzadas sin necesidad de desarrollar algoritmos desde cero. Además, la capacidad de Watson para integrarse con otras herramientas y sistemas facilita su adopción en diferentes entornos tecnológicos, permitiendo a las empresas aprovechar la IA sin realizar cambios drásticos en su infraestructura existente.

Para la toma de decisiones, IBM Watson es invaluable. Puede analizar grandes volúmenes de datos no estructurados y estructurados, proporcionando investigación de mercados accionables que ayudan a las empresas a tomar decisiones más informadas y estratégicas. Watson puede analizar registros médicos y literatura científica para ayudar a los médicos a diagnosticar enfermedades y recomendar tratamientos. En el sector financiero, puede analizar datos de mercado y patrones de comportamiento para prever tendencias y detectar fraudes. Esta capacidad de transformar datos en conocimiento útil hace que IBM Watson sea una herramienta esencial para la toma de decisiones basadas en datos.

Big Data: Apache Hadoop

Justificación: Apache Hadoop es una de las herramientas más conocidas y utilizadas en el ámbito del Big Data. Es un marco de código abierto que permite el procesamiento distribuido de grandes conjuntos de datos a través de clústeres de computadoras. Hadoop se basa en dos componentes principales: el Sistema de Archivos Distribuidos de Hadoop (HDFS) y el modelo de programación MapReduce. HDFS permite almacenar grandes cantidades de datos de manera distribuida y segura, mientras que MapReduce facilita el procesamiento paralelo de esos datos, permitiendo realizar cálculos complejos de manera eficiente.

Una de las mayores ventajas de Hadoop es su capacidad para escalar desde un solo servidor hasta miles de máquinas, cada una ofreciendo almacenamiento y procesamiento local. Esto lo hace ideal para organizaciones que manejan grandes volúmenes de datos y necesitan soluciones que puedan crecer junto con sus necesidades. Además, Hadoop es altamente flexible y puede integrarse con una variedad de herramientas y sistemas, lo que facilita su adopción en diferentes entornos empresariales.

Hadoop es invaluable porque permite a las organizaciones procesar y analizar grandes cantidades de datos rápidamente. Esto es crucial en industrias como la banca, la atención médica y el comercio minorista, donde los datos se generan a gran escala y deben ser analizados de manera eficiente. Por ejemplo, en el comercio minorista, Hadoop puede analizar datos de transacciones en tiempo real para identificar tendencias de ventas y gestionar el inventario de manera más efectiva. En el sector financiero, puede utilizarse para detectar fraudes analizando patrones inusuales en las transacciones. La capacidad de Hadoop para manejar y procesar grandes volúmenes de datos de manera rápida y eficiente lo convierte en una herramienta esencial para la toma de decisiones estratégicas y operativas.

Machine Learning: XGBoost

Justificación: XGBoost es una herramienta de Machine Learning que ha ganado una notable popularidad debido a su eficiencia y rendimiento. Es un algoritmo de boosting de gradiente, lo que significa que se enfoca en mejorar los errores de predicción de un conjunto de modelos más simples, combinándolos para crear un modelo final más preciso. XGBoost se destaca por su capacidad para manejar datos faltantes y realizar una selección automática de características, lo que lo hace extremadamente robusto y adaptable a diversas tareas de predicción.

Una de las mayores ventajas de XGBoost es su rapidez y eficiencia en términos de uso de memoria, lo que lo hace ideal para trabajar con grandes conjuntos de datos. Esto es particularmente útil en entornos académicos y de investigación donde los datos pueden ser extensos y complejos. La capacidad de XGBoost para realizar una optimización del modelo y su precisión predictiva superior lo hacen una herramienta esencial para científicos de datos y analistas que buscan obtener resultados significativos y rápidos.

En términos de toma de decisiones, XGBoost ofrece un valor significativo. Permite a las organizaciones identificar patrones y tendencias en los datos, lo que proporciona una base sólida para decisiones

informadas. Por ejemplo, en el sector financiero, XGBoost puede utilizarse para predecir el riesgo crediticio, identificando clientes con alta probabilidad de incumplimiento de pagos. En marketing, puede ayudar a segmentar clientes y personalizar campañas promocionales, mejorando la efectividad y el retorno de la inversión. La capacidad de XGBoost para transformar grandes volúmenes de datos en insights precisos y accionables lo convierte en una herramienta esencial para la toma de decisiones estratégicas y operativas.

Data Mining: WEKA

Justificación: WEKA es una herramienta de minería de datos ampliamente utilizada en el ámbito académico y profesional. Desarrollada por la Universidad de Waikato en Nueva Zelanda, WEKA proporciona una colección de algoritmos de Machine Learning y herramientas de visualización que facilitan el análisis y la exploración de datos. Una de las principales ventajas de WEKA es su interfaz gráfica de usuario intuitiva, que permite a los usuarios realizar experimentos de minería de datos sin necesidad de conocimientos profundos en programación. Esto la hace especialmente útil para estudiantes y principiantes en el campo.

WEKA soporta una variedad de tareas de minería de datos, incluyendo clasificación, regresión, clustering y reglas de asociación. Su flexibilidad y facilidad de uso permiten a los usuarios probar diferentes enfoques y técnicas rápidamente, lo que es crucial en entornos de investigación y desarrollo. Además, WEKA puede integrarse con otros sistemas y bases de datos, lo que facilita la importación y exportación de datos para su análisis.

En términos de toma de decisiones, WEKA es una herramienta valiosa porque permite a los usuarios descubrir patrones ocultos en los datos y obtener insights significativos. Por ejemplo, en el sector de la salud, WEKA puede ayudar a identificar factores de riesgo asociados con enfermedades crónicas mediante el análisis de grandes conjuntos de datos médicos. En el comercio minorista, puede usarse para analizar el comportamiento de compra de los clientes y optimizar el inventario y las estrategias de marketing. La capacidad de WEKA para proporcionar resultados precisos y fáciles de interpretar lo convierte en una herramienta esencial para la toma de decisiones basadas en datos.

Bibliografía

NexusAdmistraIntegra. (2020, January 17). *Inteligencia Artificial (IA): Ventajas y Desventajas de su Uso*. Nexus Integra. https://nexusintegra.io/es/ventajas-y-desventajas-de-la-inteligencia-artificial/#:~:text=La%20Inteligencia%20artificial%20permite%20que,autom%C3%A1tica%20y%20sin%20intervenci%C3%B3n%20humana.&text=La%20IA%20libera%20a%20las,tiempo%20a%20desarrollar%20funciones%20creativas.

Romanos, J. (2019, February 12). *Qué asistente de voz es mejor: Alexa, Siri o Google Assistant*. ADSLZone; ADSLZone. https://www.adslzone.net/reportajes/domotica/google-assistant-alexa-siri

Conocimiento, del. (2016, June 28). *Las 7 V del Big data: Características más importantes - IIC*. Instituto de Ingeniería Del Conocimiento. https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/

Todo lo que necesitas saber sobre Inteligencia Artificial. (2016). Bismart.com. https://landing.bismart.com/inteligencia-artificial

¿Qué es la inteligencia artificial (IA)? (2014). Oracle.com. https://www.oracle.com/mx/artificial-intelligence/what-is-ai/

Ortega, C. (2023, November 7). *Modelos de machine learning: Qué son, tipos y aplicaciones*. QuestionPro. https://www.questionpro.com/blog/es/modelos-de-machine-learning/

Bello, E. (2023, October 31). *¿Qué es el minado de Datos o Data Mininig? Técnicas y pasos a seguir*. Thinking for Innovation. https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/

Ortega, C. (2023, October). *Herramientas de inteligencia artificial: 5 ejemplos y características*. QuestionPro. https://www.questionpro.com/blog/es/herramientas-de-inteligencia-artificial/#:~:text=Las%20herramientas%20de%20IA%20son,basadas%20en%20patrones%20y%20conocimientos.

¿Cómo programar inteligencia artificial? Lenguajes y Claves. (2023, February 20). UNIR México; UNIR México. https://mexico.unir.net/noticias/ingenieria/programar-inteligencia-artificial/

Pérez, L. (2023, June). *Descubre las herramientas y lenguajes de la IA*. Neuroflash. https://neuroflash.com/es/blog/descubre-las-herramientas-emocionantes-de-ia/#:~:text=Hay%20varios%20lenguajes%20de%20programaci%C3%B3n,bibliotecas%20de%20c%C3%B3digo%20abierto%20disponibles.

https://www.facebook.com/grokkeepcoding. (2021, November 25). *Top 5 Lenguajes del Big Data | KeepCoding Bootcamps*. KeepCoding Bootcamps. https://keepcoding.io/blog/los-5-lenguajes-del-big-data/

“Data mining”, definición, ejemplos y aplicaciones - Iberdrola. (2024). Iberdrola. https://www.iberdrola.com/innovacion/data-mining-definicion-ejemplos-y-aplicaciones#:~:text=%C2%BFQU%C3%89%20ES%20EL%20'DATA%20MINING,sentido%20y%20convertirla%20en%20conocimiento.

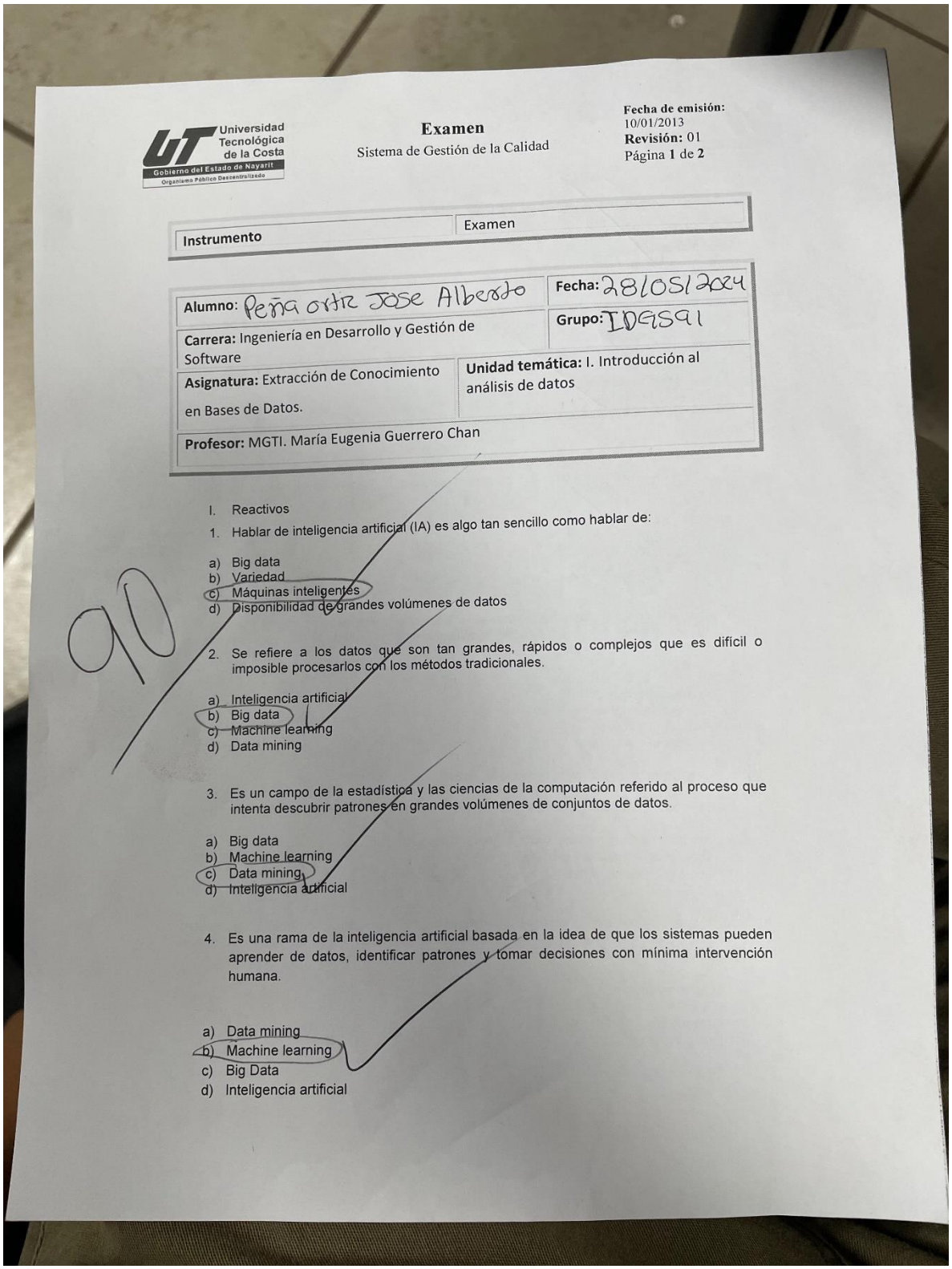
Bello, E. (2023, July 31). *Mejores herramientas de Machine Learning 2024*. Thinking for Innovation. https://www.iebschool.com/blog/herramientas-business-intelligence-big-data/#:~:text=Las%20herramientas%20de%20Machine%20Learning,la%20construcci%C3%B3n%20de%20modelos%20anal%C3%ADticos.

“Machine Learning”: definición, tipos y aplicaciones prácticas - Iberdrola. (2024). Iberdrola. https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico

Conocimiento, del. (2016, October 13). *7 Herramientas Big Data para tu empresa - IIC*. Instituto de Ingeniería Del Conocimiento. https://www.iic.uam.es/innovacion/herramientas-big-data-para-empresa/

¿Qué es la minería de datos? | Definición, importancia y tipos | SAP Insights. (2017). SAP. https://www.sap.com/latinamerica/products/technology-platform/hana/what-is-data-mining.html#:~:text=miner%C3%ADa%20de%20datos%3F-,Data%20Mining%20es%20el%20proceso%20de%20uso%20de%20herramientas%20anal%C3%ADticas,sistemas%20aprendan%20de%20la%20experiencia.

Examen Unidad 1



Universidad Tecnológica de la Costa
Gobierno del Estado de Nayarit
Organismo Público Descentralizado

Examen
Sistema de Gestión de la Calidad

Fecha de emisión: 10/01/2013
Revisión: 01
Página **1** de **2**

Instrumento	Examen

Alumno: Peña ortiz Jose Alberto	**Fecha:** 28/05/2024
Carrera: Ingeniería en Desarrollo y Gestión de Software	**Grupo:** IDGS91
Asignatura: Extracción de Conocimiento en Bases de Datos.	**Unidad temática:** I. Introducción al análisis de datos
Profesor: MGTI. María Eugenia Guerrero Chan	

I. Reactivos

1. Hablar de inteligencia artificial (IA) es algo tan sencillo como hablar de:

a) Big data
b) Variedad
c) Máquinas inteligentes
d) Disponibilidad de grandes volúmenes de datos

90

2. Se refiere a los datos que son tan grandes, rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales.

a) Inteligencia artificial
b) Big data
c) Machine learning
d) Data mining

3. Es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.

a) Big data
b) Machine learning
c) Data mining
d) Inteligencia artificial

4. Es una rama de la inteligencia artificial basada en la idea de que los sistemas pueden aprender de datos, identificar patrones y tomar decisiones con mínima intervención humana.

a) Data mining
b) Machine learning
c) Big Data
d) Inteligencia artificial

Unidad 2: Preparación de los datos

Instalación de Lenguaje R y RStudio

Instrumento	*Práctica de ejercicios*

Alumno: Peña Ortiz Jose Alberto	**Fecha:** 30 de mayo de 2024
Carrera: IDGS	**Grupo:** IDGS91
Asignatura: Extracción de Conocimiento en Base de Datos	**Unidad temática:** Unidad 2: Preparación de los datos
Profesor: María Eugenia Guerrero Chan	

Contenido

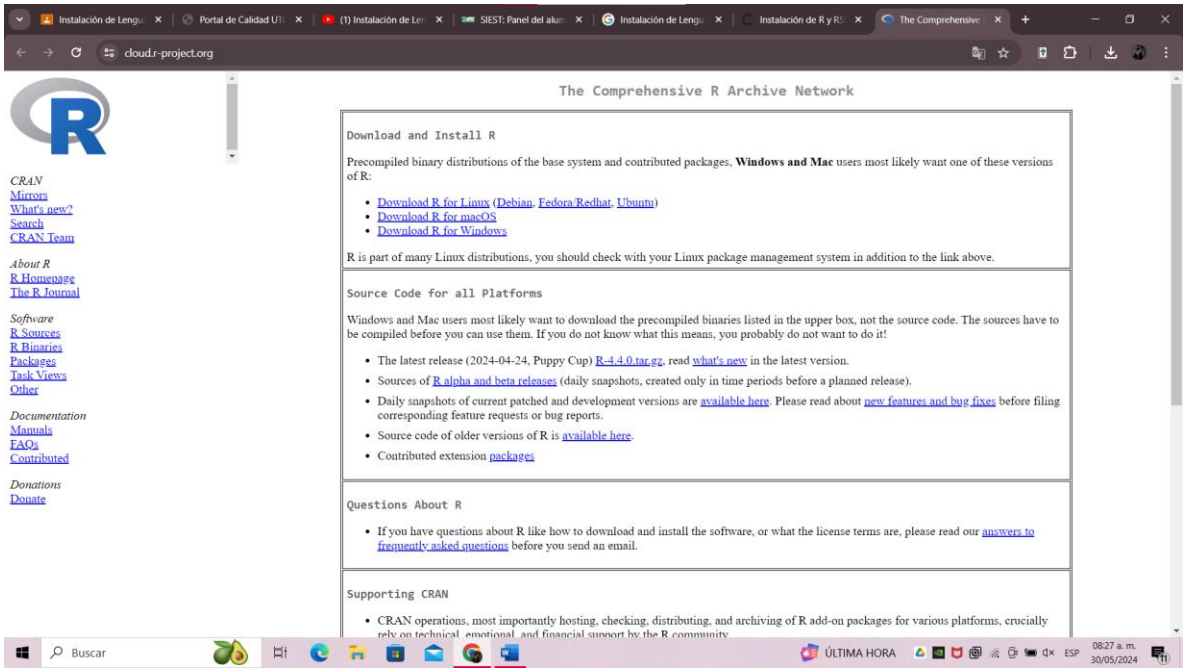
I.- Ejercicios a resolver:

Instalación de Lenguaje R y RStudio

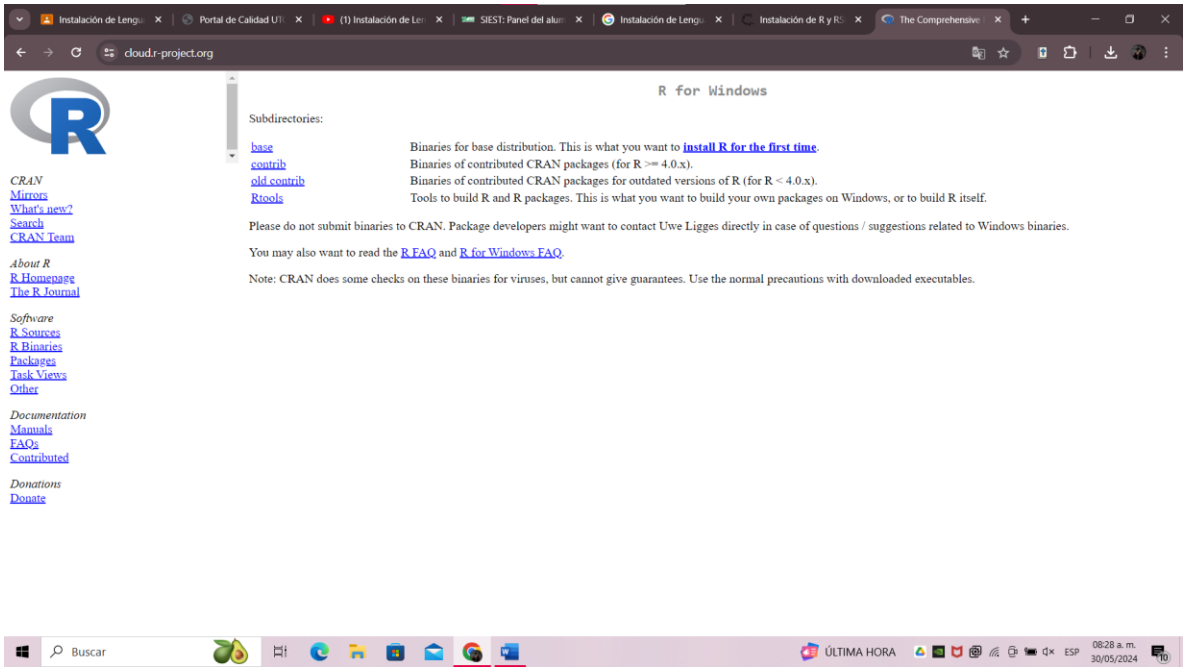
II.-Procedimientos y resultados:

Instalación de R en Windows

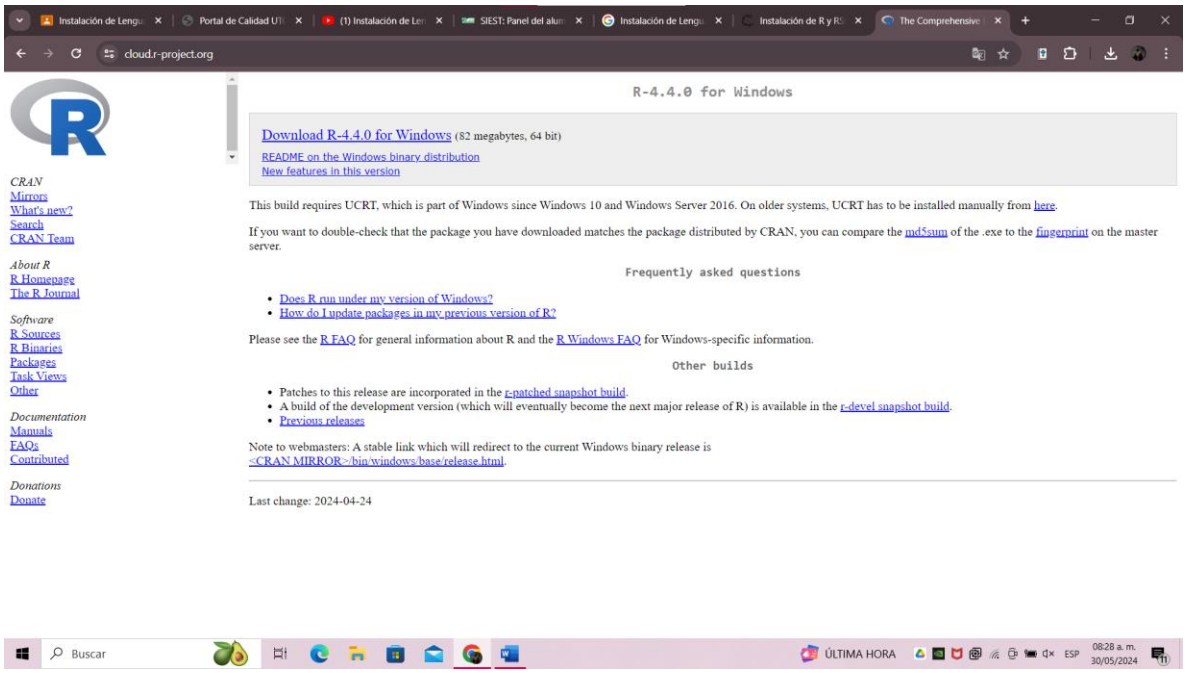
Para descargar R, primero necesitas ir a CRAN (Comprehensive R Archive Network), que es el repositorio principal de software de R. CRAN está formado por varios servidores distribuidos por todo el mundo, conocidos como espejos, que ayudan a distribuir tanto R como sus paquetes. Nosotros utilizaremos el servidor en la nube, por lo que no tendrás que elegir un servidor cercano, ya que se detectará automáticamente. Para empezar, abre tu navegador favorito, dirígete al sitio oficial del servidor en la nube de R y haz clic en el enlace "Download R for Windows".

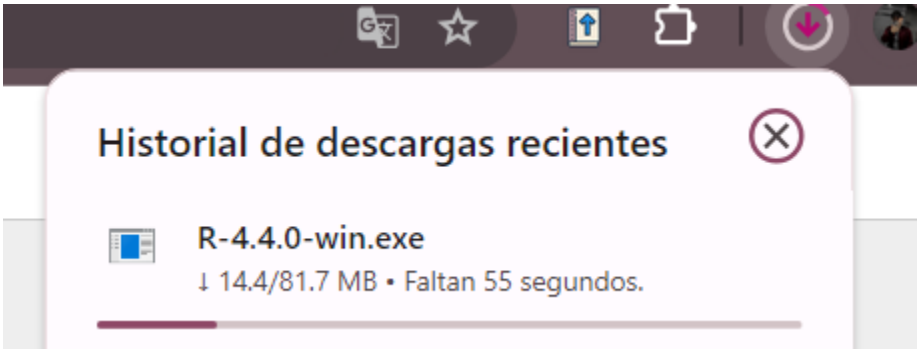


En la página que se abre, selecciona el enlace "install R for the first time" que se encuentra en la parte superior.

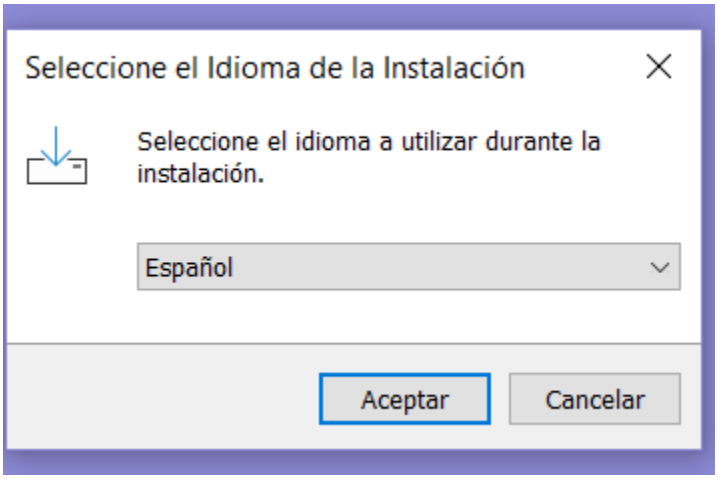


Luego, haz clic en "Download R 4.4.0 for Windows". El número que aparece después de R indica la versión que vas a instalar, y puede variar con las actualizaciones. Solo asegúrate de que sea una versión de R-4.

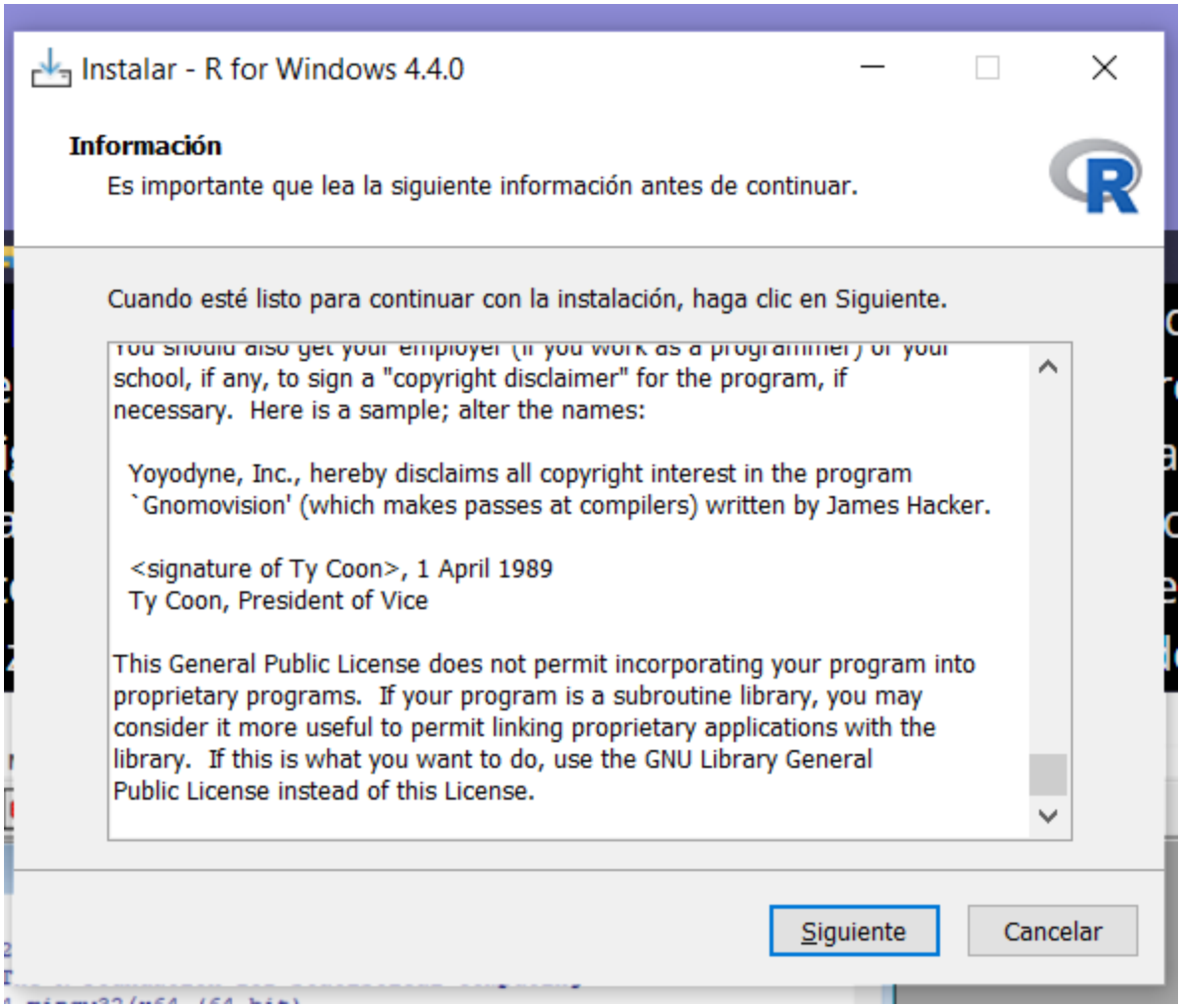




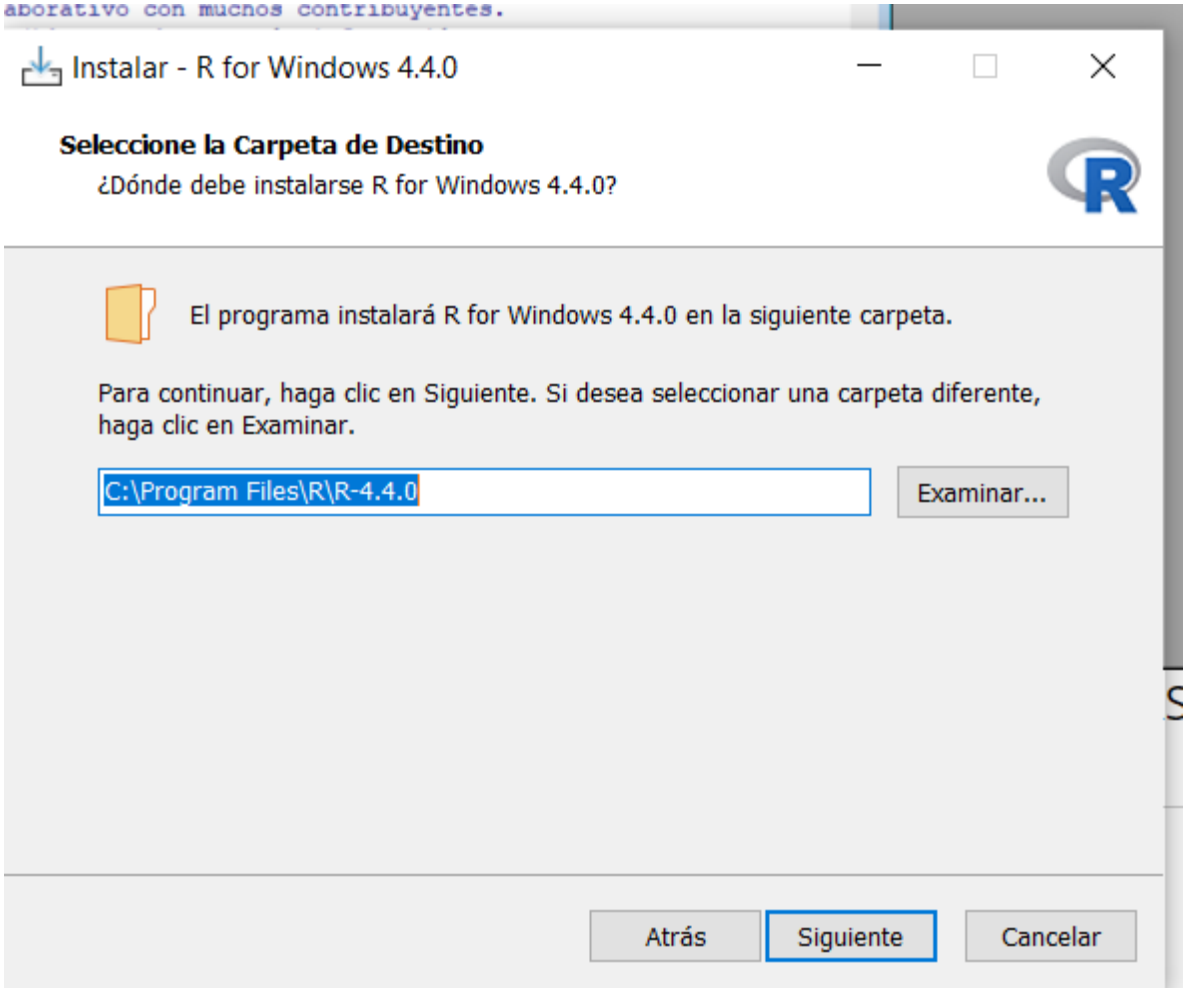
Haz doble clic en el archivo descargado para ejecutarlo.



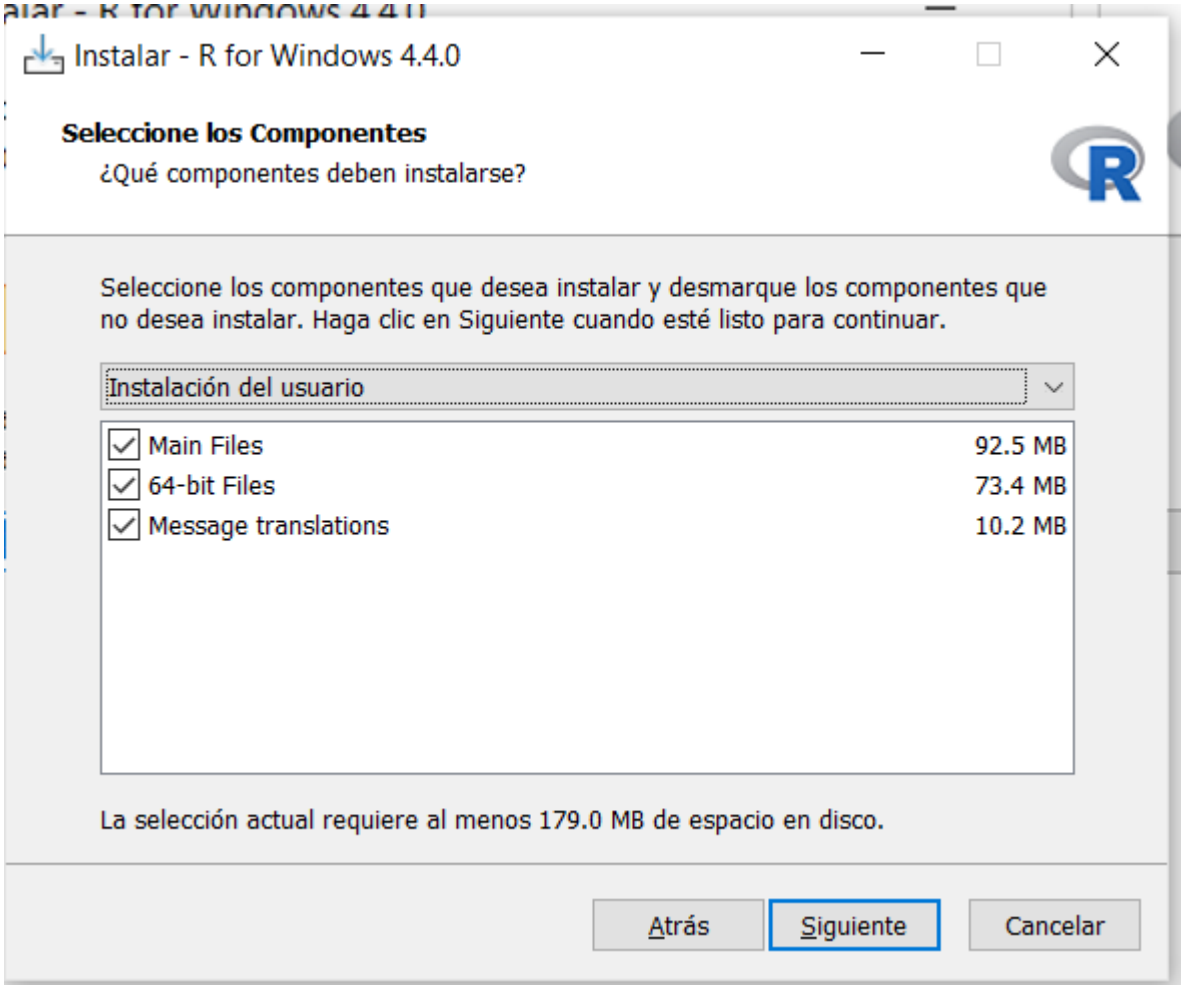
Después de dar en aceptar para seleccionar el idioma de instalación, nos mostrara la ventana de los términos y condiciones damos en siguiente.



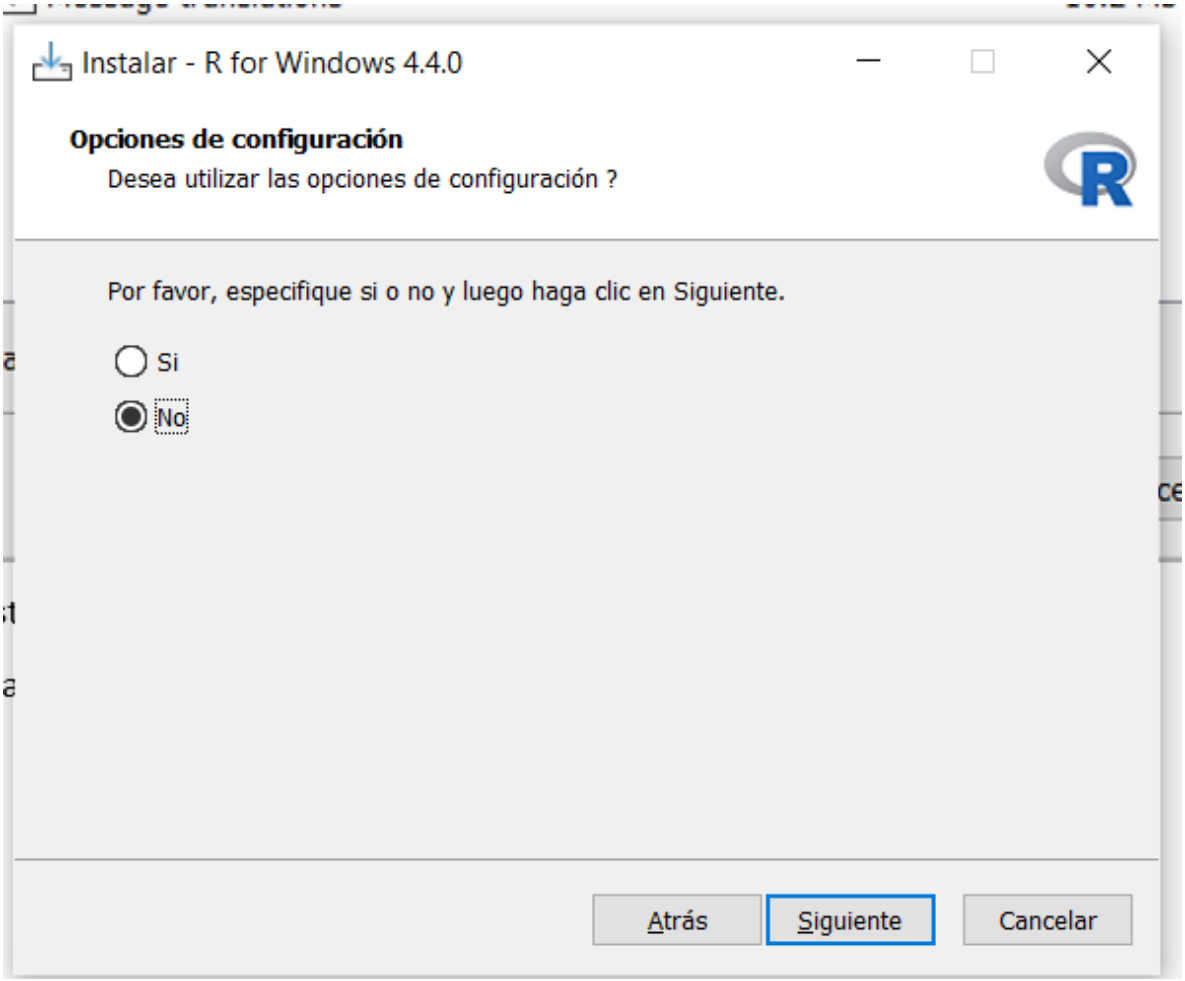
Nos mostrara una ventana donde nos pedirá la ruta donde instalaremos R, podemos dejar la que nos mostrara por defecto y darle en siguiente.



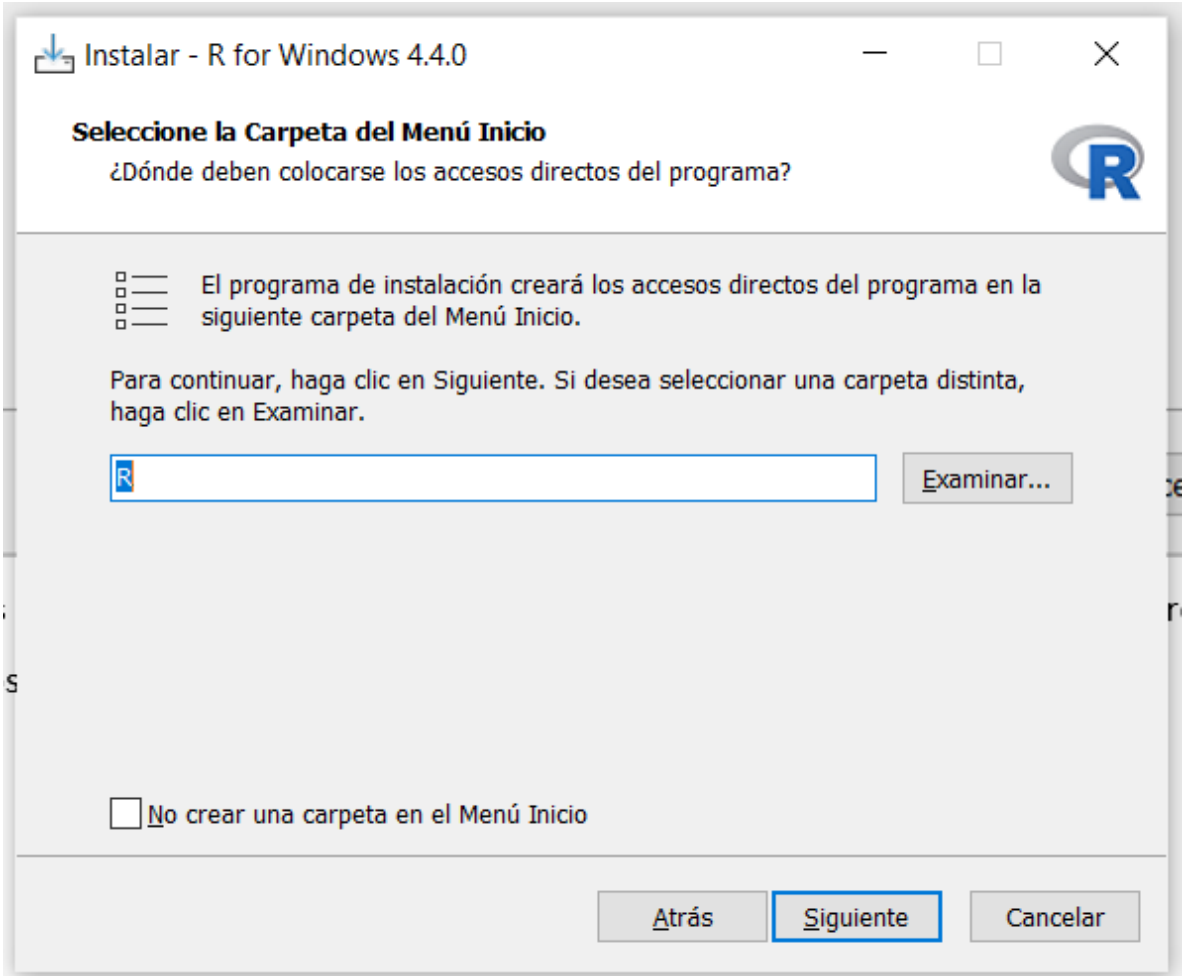
La siguiente ventana nos mostrara los componentes que se instalaran en nuestro dispositivo de cómputo, damos en siguiente.



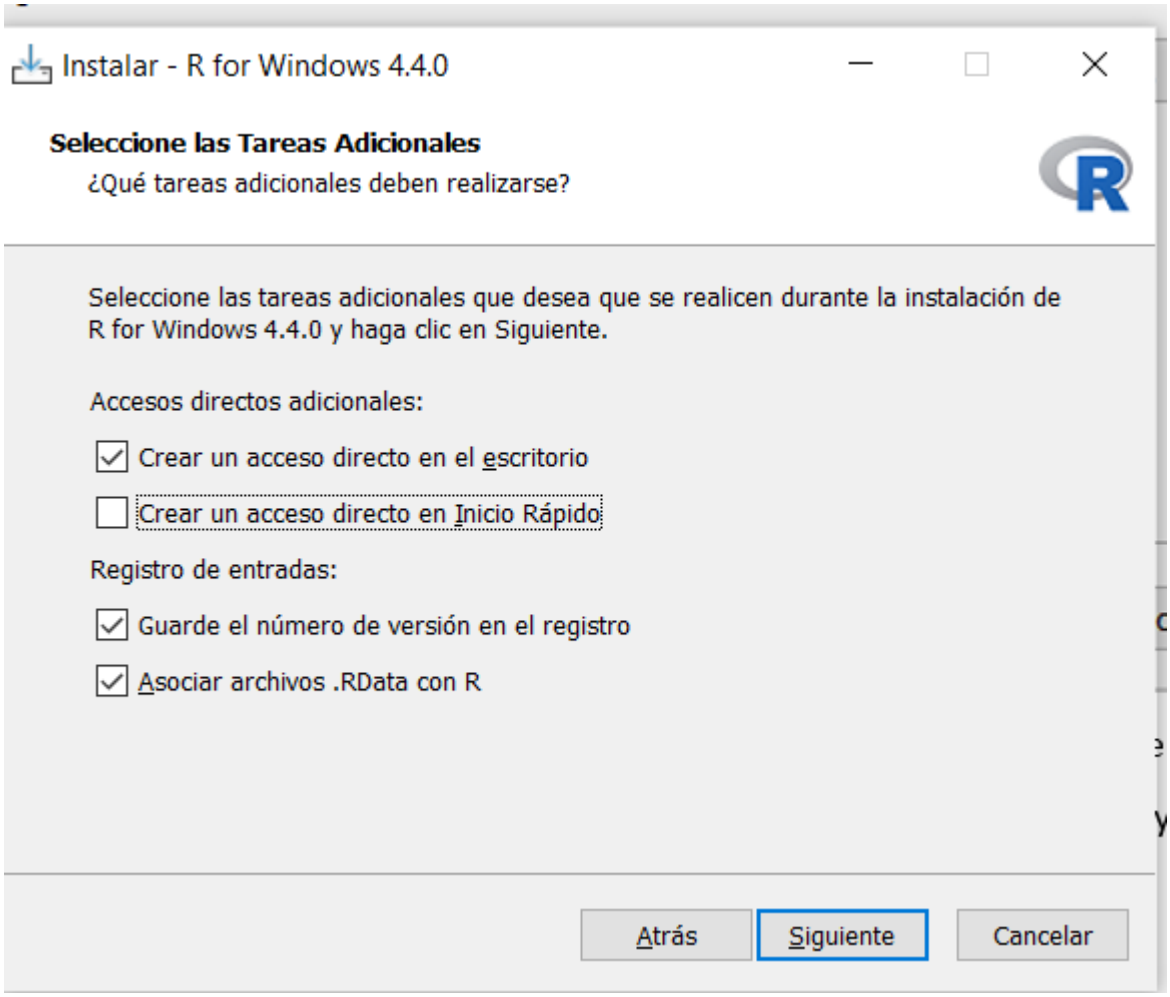
Nos mostrara una ventana donde nos preguntara si queremos las opciones de configuración, seleccionamos que “No” y damos siguientes.



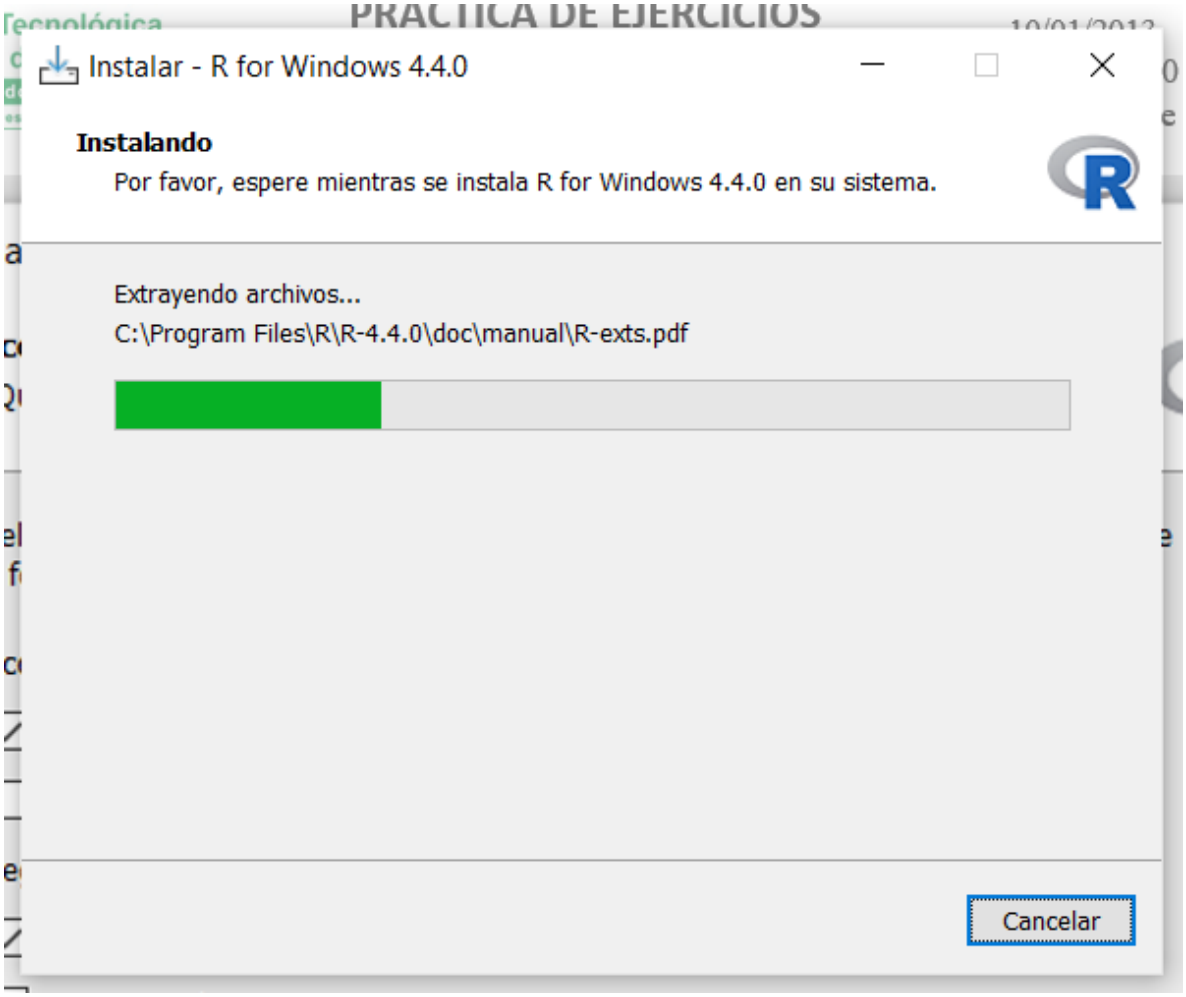
Después la siguiente ventana nos mostrara donde queremos colocar el acceso directo a R podemos dejarlo por defecto y darle en Siguiente



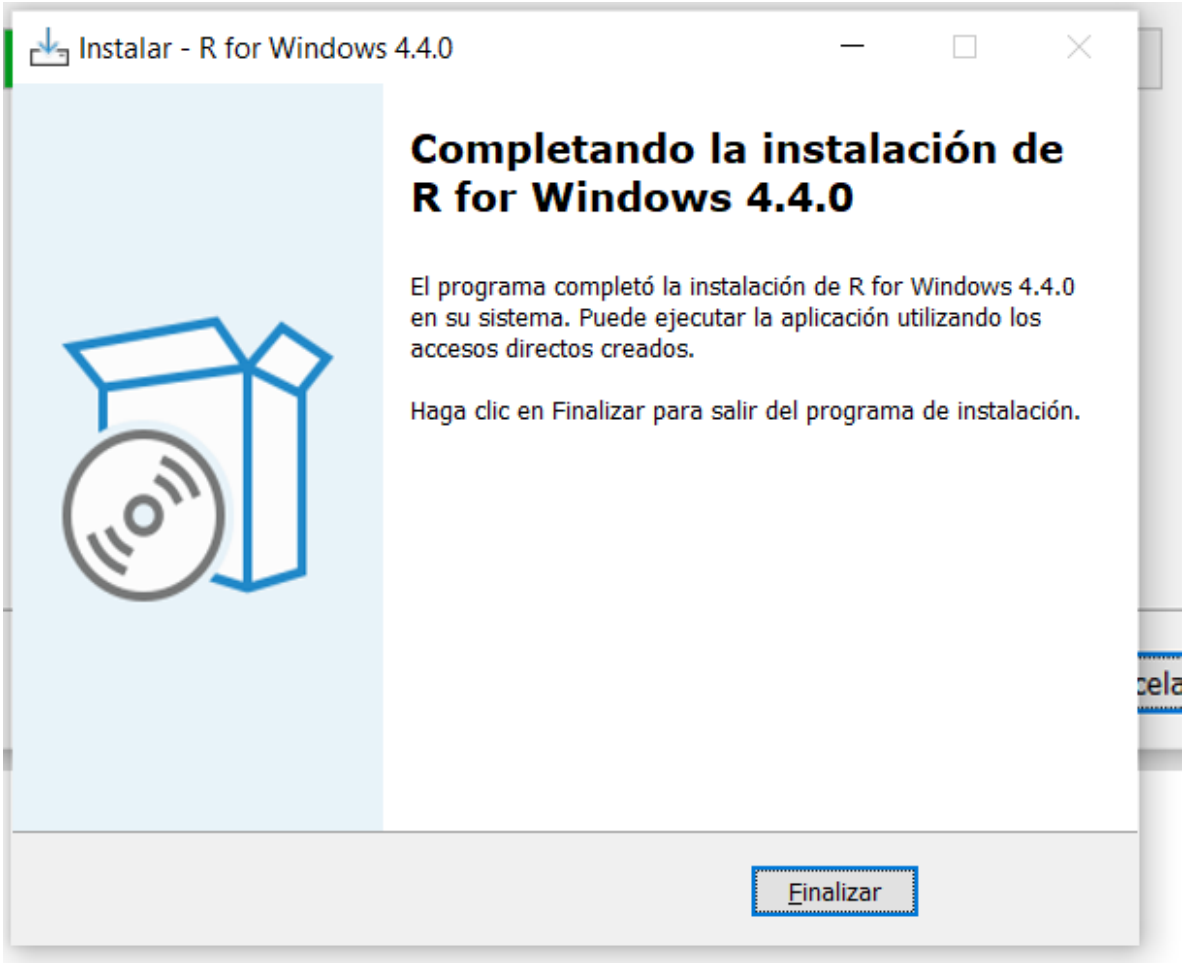
Ya después de dar siguiente, nos preguntara que si queremos seleccionar las tareas adicionales, podemos dejarlo por defecto ya que viene con casillas seleccionada y damos en siguiente.



Una vez dado siguiente comenzará la instalación de R y se mostrará una barra de carga del proceso.

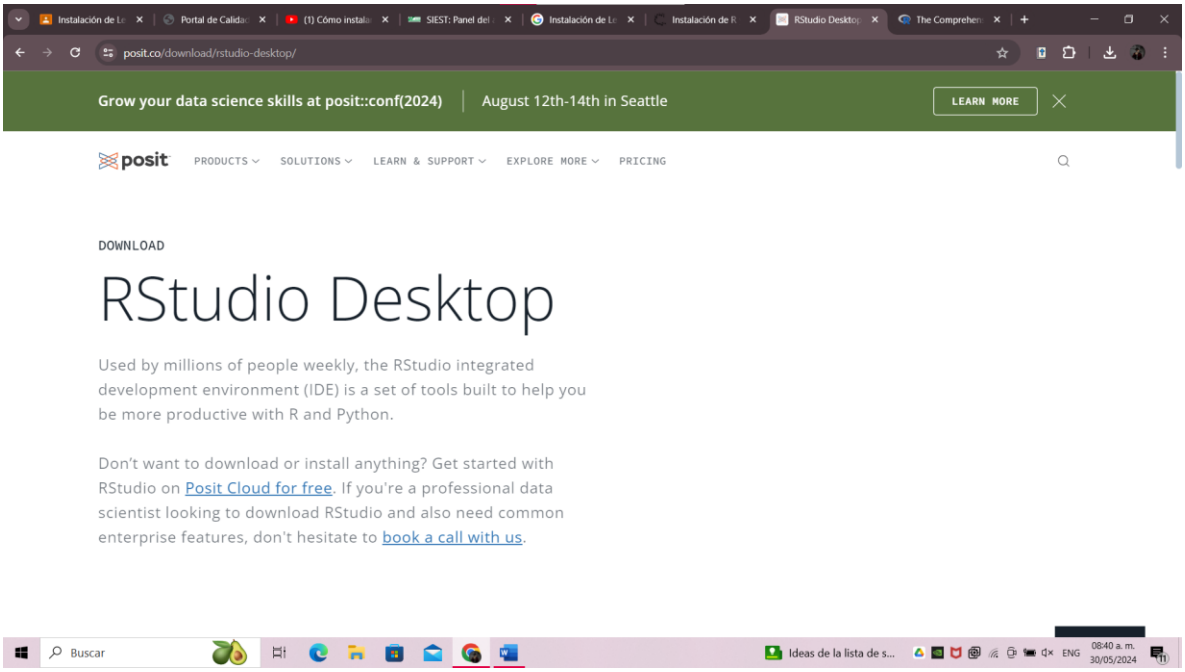


Una vez que termine de instalarse, se mostrara una ventana donde nos indicara que se Completó la instalación de R para Windows



Instalación de RStudio en Windows

Abre de nuevo tu navegador y dirígete al sitio oficial de RStudio.



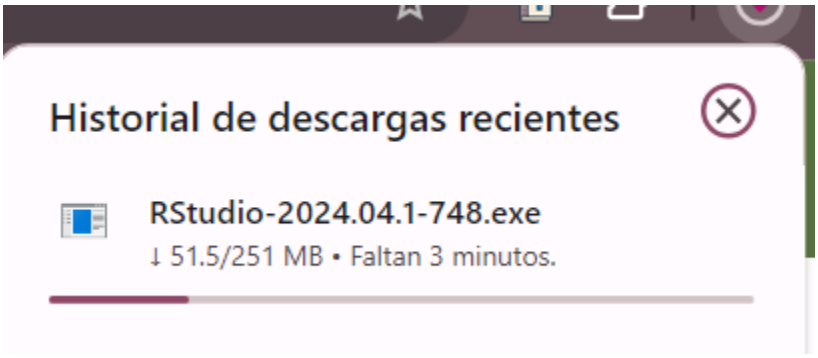
Hacemos clic en “Install RStudio DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS”.

2: Install RStudio

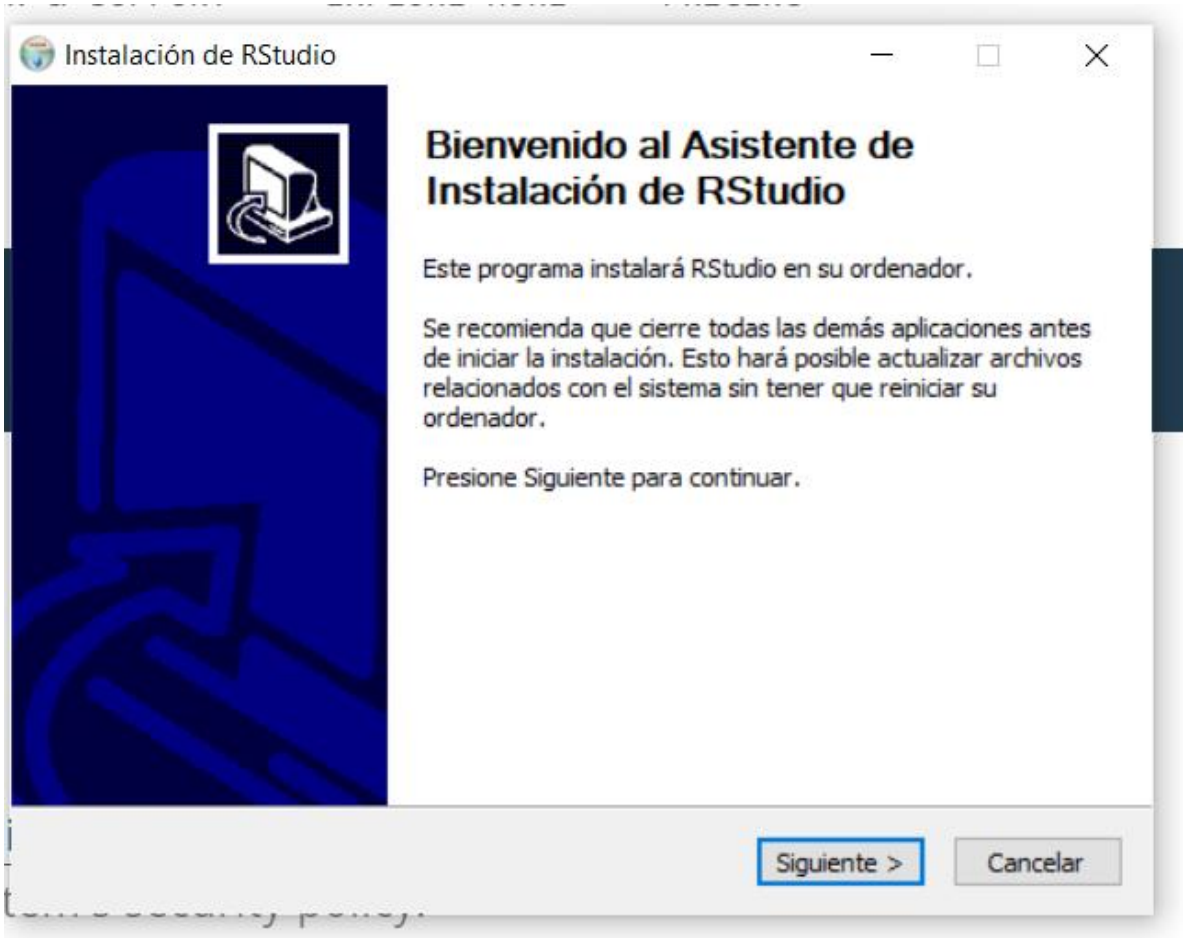
DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 263.07 MB | SHA-256: 44C8797C | Version: 2024.04.1+748 | Released: 2024-05-11

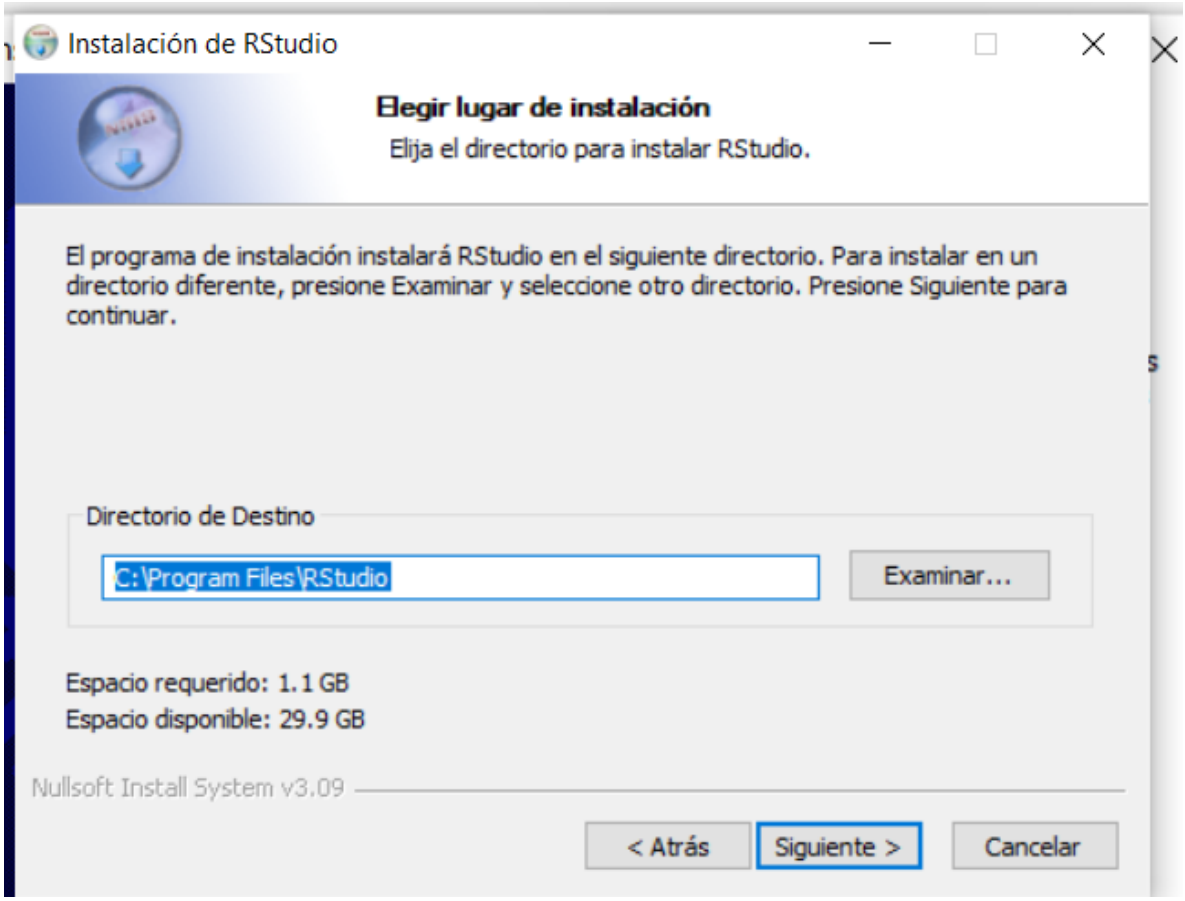
Comenzará la descargar del instalador de RStudio



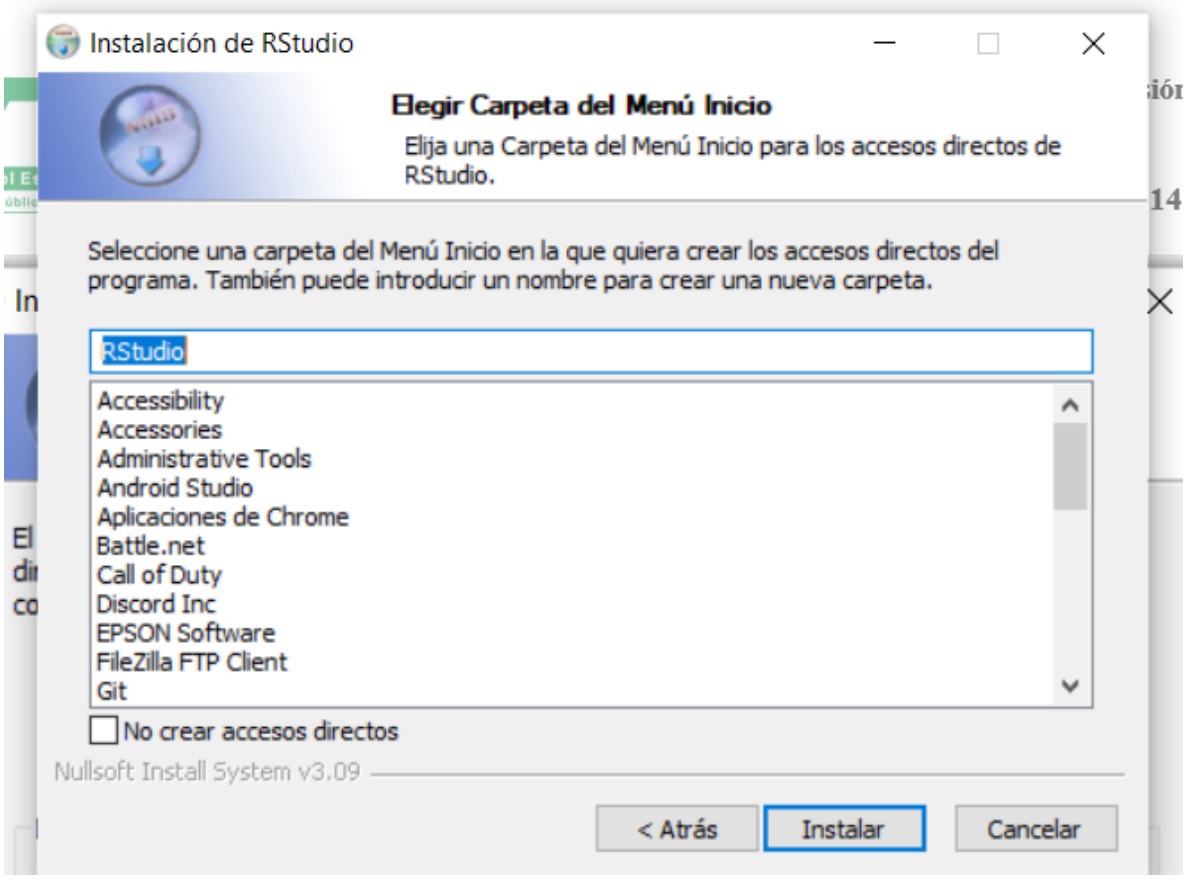
Una vez finalizada la descarga damos clic en el ejecutable de RStudio y nos mostrara la ventana de "Bienvenido al asistente de instalación de RStudio" damos en siguiente.



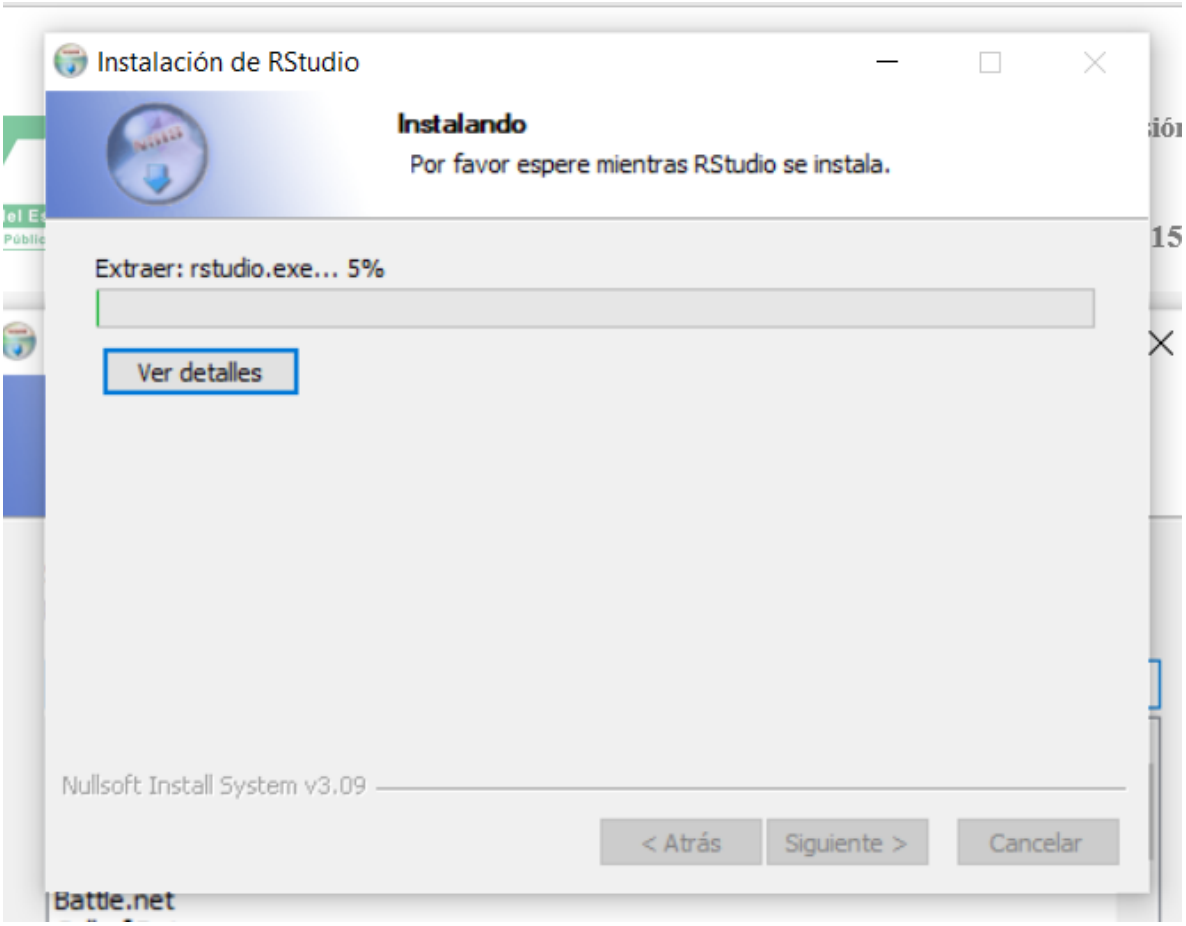
Nos preguntara donde queremos instalar RStudio, podemos dejar la ruta que viene por defecto y darle en siguiente



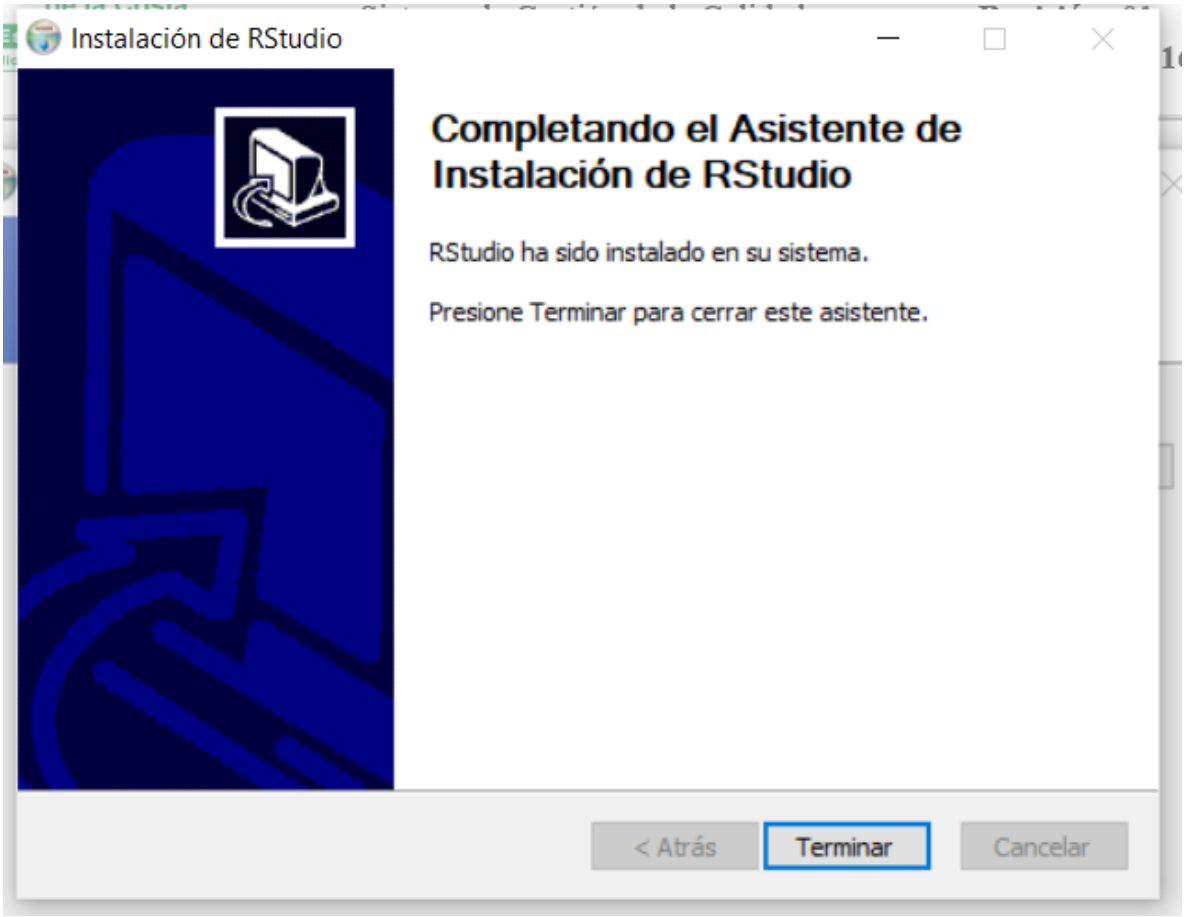
Nos pedirá elegir una carpeta pero como en el caso anterior podemos dejarlo por defecto como lo muestra la ventana y darle en instalar.



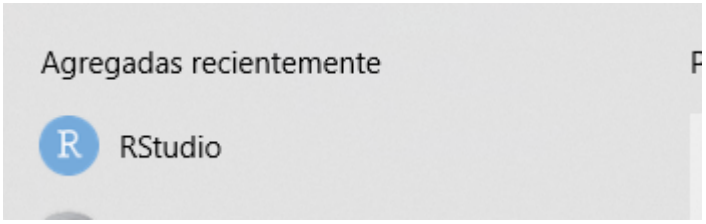
Después de dar clic en instalar empezara la descarga e instalación de RStudio a tu dispositivo de cómputo mostrando en una barra de progreso.



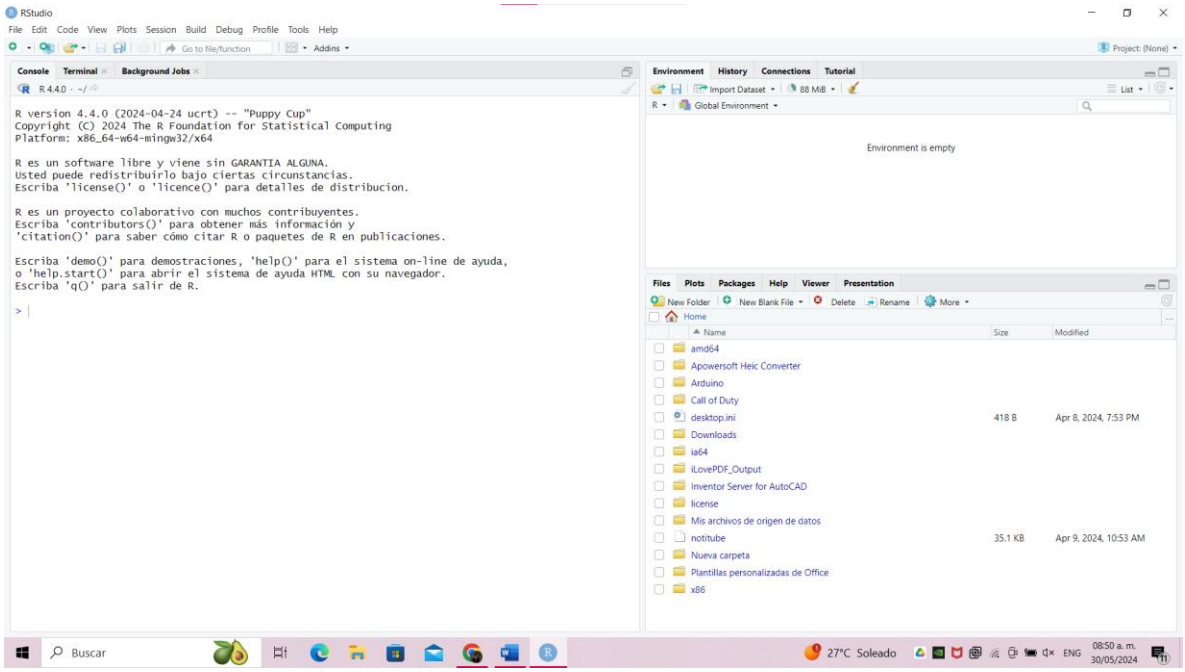
Una vez finalizada la instalación, nos saldrá una última ventana que nos dirá que se completó correctamente la instalación de RStudio, damos en terminar.



Podemos buscar RStudio en nuestras aplicaciones instaladas con este nombre e icono como se muestra en la siguiente imagen



Damos clic en la aplicación y se abrirá el programa mostrando esta interfaz.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console
Terminal
Background Jobs
R 4.4.0 · ~/
R version 4.4.0 (2024-04-24 ucrt) -- "Puppy Cup"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64
R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.
R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.
>
Environment
History
Connections
Tutorial
Import Dataset
88 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
New Folder
New Blank File
Delete
Rename
More
Home
Name
Size
Modified
amd64
Apowersoft Heic Converter
Arduino
Call of Duty
desktop.ini
418 B
Apr 8, 2024, 7:53 PM
Downloads
ia64
iLovePDF_Output
Inventor Server for AutoCAD
license
Mis archivos de origen de datos
notitube
35.1 KB
Apr 9, 2024, 10:53 AM
Nueva carpeta
Plantillas personalizadas de Office
x86
Buscar
27°C Soleado
ENG
08:50 a. m.
30/05/2024

Bibliografía

The Comprehensive R Archive Network. (2024). R-Project.org. https://cloud.r-project.org/

Posit. (2024, May 7). Posit. https://posit.co/download/rstudio-desktop/

Mapa conceptual_U2

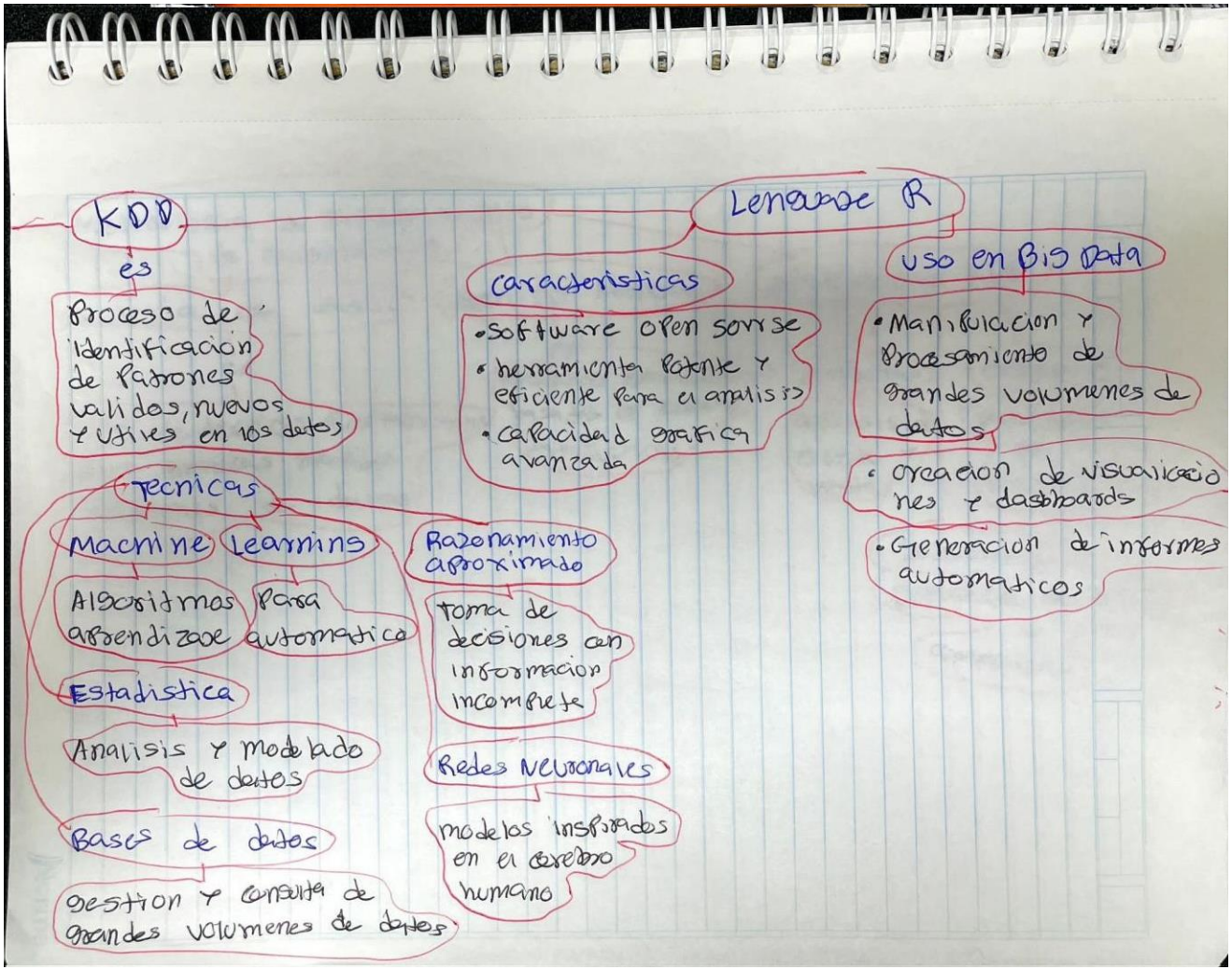
Instrumento	*Mapa conceptual o Cuadro sinóptico*

Alumno: Peña Ortiz Jose Alberto	**Fecha: O6/O6/2024**
Carrera: IDSG	**Grupo: IDGS91**
Asignatura: Extracción de Conocimiento en Base de Datos	**Unidad temática: Unidad 2: Preparación de los datos**
Profesor: MGTI María Eugenia Guerrero Chan	

Realiza un mapa conceptual con la información que expusieron los equipos en clase y también te puedes apoyar con la información que les subí en la plataforma de classroom.

NOTA: El mapa conceptual lo realizarán a mano en una hoja blanca o de libreta, una vez hecho tomarle foto y ponerlo en el formato de mapa conceptual.

Mapa conceptual_U2


KDD
es
Proceso de identificacion de Patrones validos, nuevos y utiles en los datos
Tecnicas
Machine Learning
Algoritmos aprendizaje
Para automatica
Estadistica
Analisis y modelado de datos
Bases de datos
gestion y consulta de grandes volumenes de datos
Razonamiento aproximado
Toma de decisiones con informacion incompleta
Redes Neuronales
modelos inspirados en el cerebro humano
Lenguaje R
caracteristicas
• Software open sorrse
• herramienta Potente y eficiente para el analisis
• capacidad grafica avanzada
Uso en Big Data
• Manipulacion y Procesamiento de grandes volumenes de datos
• creacion de visualizaciones y dashboards
• Generacion de informes automaticos

Bibliografía

ECBD_U2_clase.pdf. (2019). *ECBD_U2_clase.pdf*. Google Docs. https://drive.google.com/file/d/1M_tmEiuFlEDXEeg9xUXPPvTJRAN3EynV/view

Examen Unidad 2

Universidad Tecnológica de la Costa
Gobierno del Estado de Nayarit
Organismo Público Descentralizado

Examen
Sistema de Gestión de la Calidad

Fecha de emisión: 10/01/2013
Revisión: 01
Página **1** de **2**

Instrumento	Examen

Alumno: Peña Ortiz Jose Alberto	**Fecha:** 07/Junio/24
Carrera: Ingeniería en Desarrollo y Gestión de Software	**Grupo:** IDGS
Asignatura: Extracción de Conocimiento en Bases de Datos.	**Unidad temática:** II. Preparación de los datos
Profesor: MGTI. María Eugenia Guerrero Chan	

I. Reactivos

1. En el análisis de datos, este tipo de datos son los relativos a las "cualidades", este tipo de información relacionada con los adjetivos.

a) Cuantitativos b) Biométricos c) Cualitativos d) Relativos

2. La estatura, el peso, etc. Son ejemplos de tipos de datos:

a) Datos de transacciones b) Discretas c) Continuas d) Binarios

3. Un _____se encarga de extraer datos de las bases de datos operacionales o fuentes externa, transformar, consolidar, integrar, chequear la integridad y centralizar los datos que la empresa genera en su actividad diaria de negocios y/o información externa con la que esté relacionada.

a) Dato semiestructurado
b) Dato estructurado
c) Data Warehouse
d) Análisis predictivo

4. Este tipo de dato se encuentran organizados mediante una serie de filas y columnas bien definidas. Son los que se usan de manera habitual en la mayor parte de las bases de datos relacionales

a) Datos no estructurados
b) Datos estructurados
c) Biométricos
d) Datos semiestructurados

5. Es también conocida como depuración de datos, es el primer paso en el proceso de preparación de datos. Implica identificar errores en un conjunto de datos y corregirlos para garantizar que solo se transfieran datos limpios y de alta calidad a los sistemas de destino.

a) Inteligencia artificial
b) Análisis
c) Data Warehouse
d) Ninguna de las anteriores

6. Son tipos de datos que incluye escaneo de la retina, huellas digitales, reconocimiento genético o facial, etc.

a) Datos de transacciones
b) Web
c) Biométricos
d) Ninguna de las anteriores

7. Son características de un Data Warehouse, excepto:

a) Administra grandes cantidades de información
b) Guarda histórico de datos
c) Las variables y constantes
d) Integra y asocia información de muchas fuentes

8. Son técnicas de limpieza de datos, excepto:

a) Data cleaning b) Data transformation c) Data integration d)Data exploration

9. Es el proceso de extraer datos capturados dentro de fuentes semiestructuradas y no estructuradas, como correos electrónicos, documentos PDF, formularios PDF, archivos de texto, redes sociales, códigos de barras e imágenes.

a) Extracción de datos
b) Transformación de datos
c) La carga de datos
d) Ninguna de las anteriores

10. Es el proceso de compilación de datos a partir de un número ilimitado de fuentes, su posterior organización y centralización en un único repositorio.

a) Limpieza de datos
b) Proceso ETL
c) Data Warehouse
d) Ninguna de las anteriores

Unidad 3: Análisis supervisado

Practica ejercicios 1 unidad 3

Instrumento	*Práctica de ejercicios*

Alumno: Peña Ortiz Jose Alberto	**Fecha:** 20 de junio de 2024
Carrera: Ingeniería en Desarrollo y Gestión de Software	**Grupo:** IDGS91
Asignatura: Extracción de Conocimiento en Bases de Datos.	**Unidad temática:** II. Preparación de los Datos
Profesor: MGTI. María Eugenia Guerrero Chan	

Contenido

I.- Ejercicios a resolver:

Instrucciones:

Realiza lo siguiente en el programa de RStudio.

1.- Crea una matriz y titula las columnas y filas, así como se indica en el siguiente ejemplo e ingrese cantidades para las ventas:

	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
David	452	456	945	792
Esteba	159	573	453	597
Jose	593	571	579	264
Fabricio	351	562	542	513
Rafael	586	596	579	195

Cálculos

1.- Obtener la venta mayor por trimestre.

2.-Obtener la venta menor por trimestre.

3.- Obtener la venta promedio por vendedor.

4.- Obtener la venta promedio por trimestre.

5.- Cantidad de ventas registradas en la tabla.

6.- Mostrar la venta mayor.

7.- Mostrar la venta menor.

Nota 1: El alumno ingresará las ventas

Nota 2: Dejar evidencia en cada uno de los pasos y el resultado final.

II.-Procedimientos y resultados: (Poner aquí la estructura y orden de la información)

Creación de la matriz

```
# Crear la nueva matriz de ventas
ventas <- matrix(c(452, 159, 593, 351, 586,456, 573, 571, 562, 596,945, 453, 579, 542, 579,79
rownames(ventas) <-c("David", "Esteba", "Jose", "Fabricio", "Rafael")
colnames(ventas) <- c("Trimestre 1", "Trimestre 2", "Trimestre 3", "Trimestre 4")
ventas
```

1.- Obtener la venta mayor por trimestre.

```
# 1. Obtener la venta mayor por trimestre
venta_mayor_trimestre <- apply(ventas, 2, max)
venta_mayor_trimestre
```

```
Trimestre 1 Trimestre 2 Trimestre 3 Trimestre 4
        597         596         945         792
>
```

2.-Obtener la venta menor por trimestre.

```
# 2. Obtener la venta menor por trimestre
venta_menor_trimestre <- apply(ventas, 2, min)
venta_menor_trimestre
```

```
Trimestre 1 Trimestre 2 Trimestre 3 Trimestre 4
        452         159         513         195
>
```

3.- Obtener la venta promedio por vendedor.

```
# 3. Obtener la venta promedio por vendedor
venta_promedio_vendedor <- apply(ventas, 1, mean)
venta_promedio_vendedor
```

```
> venta_promedio_vendedor
   David   Esteba     Jose Fabricio   Rafael
  388.75   546.50   639.00   623.00   392.25
```

4.- Obtener la venta promedio por trimestre.

```r
# 4. Obtener la venta promedio por trimestre
venta_promedio_trimestre <- apply(ventas, 2, mean)
venta_promedio_trimestre
```

```
> venta_promedio_trimestre
Trimestre 1 Trimestre 2 Trimestre 3 Trimestre 4
      555.2       403.4       640.6       472.4
```

5.- Cantidad de ventas registradas en la tabla.

```r
# 5. Cantidad de ventas registradas en la tabla
cantidad_ventas <- length(ventas)
cantidad_ventas
```

```
> cantidad_ventas
[1] 20
```

6.- Mostrar la venta mayor.

```r
# 6. Mostrar la venta mayor
venta_mayor <- max(ventas)
venta_mayor
```

```
> venta_mayor
[1] 945
```

7.- Mostrar la venta menor.

```r
# 7. Mostrar la venta menor
venta_menor <- min(ventas)
venta_menor
```

```
> venta_menor
[1] 159
>
```

Practica ejercicios 2 unidad 3

Instrumento	*Práctica de ejercicios*

Alumno: Peña Ortiz Jose Alberto	**Fecha:** 21 de junio de 2024
Carrera: Ingeniería en Desarrollo y Gestión de Software	**Grupo:** IDGS91
Asignatura: Extracción de Conocimiento en Bases de Datos.	**Unidad temática:** III. Análisis supervisado
Profesor: MGTI. María Eugenia Guerrero Chan	

Contenido

I.- Ejercicios a resolver:

Instrucciones:

Realiza lo siguiente en el programa de RStudio.

1.- Crea una matriz y titula las columnas y filas, así como se indica en el siguiente ejemplo e ingrese calificaciones:

	U1	U2	U3	U4
Fabian	83	86	77	93
Ramon	93	95	76	65
Luis	98	81	78	71
Liliana	68	87	92	97
Diana	86	79	81	78
Erika	95	76	83	69
Karina	75	73	67	98
Jorge	75	93	95	90

Cálculos

1.- Obtener la calificación mayor.

2.-Obtener la calificación menor.

3.- Obtener la calificación promedio por unidad.

4.- Obtener la calificación promedio por alumno.

5.- Mostrar cuantas calificaciones están dadas de alta.

6.- Obtener la mediana de las calificaciones.

Nota 1: El alumno ingresará las calificaciones.

Nota 2: Dejar evidencia en cada uno de los pasos y el resultado final.

II.-Procedimientos y resultados: (Poner aquí la estructura y orden de la información)

Creación de la matriz

```
# Crear la matriz de calificaciones
calificaciones <- c(83, 86, 77, 93,93, 95, 76, 65,98, 81, 78, 71,68, 87, 92, 97,86, 79, 81, 7

calificaciones <- matrix(calificaciones, 8, 4, byrow = T)

# Asignar nombres a las filas y columnas
rownames(calificaciones) <- c("Fabian", "Ramon", "Luis", "Liliana", "Diana", "Erika", "Karina
colnames(calificaciones) <- c("U1", "U2", "U3", "U4")
calificaciones
```

```
> calificaciones
        U1 U2 U3 U4
Fabian  83 86 77 93
Ramon   93 95 76 65
Luis    98 81 78 71
Liliana 68 87 92 97
Diana   86 79 81 78
Erika   95 76 83 69
Karina  75 73 67 98
Jorge   75 93 95 90
```

1.- Obtener la calificación mayor.

```
# 1. Obtener la calificación mayor
calif_mayor <- max(calificaciones)
calif_mayor
```

```
> calif_mayor
[1] 98
```

2.-Obtener la calificación menor.

```
# 2. Obtener la calificación menor
calif_menor <- min(calificaciones)
calif_menor
```

```
> calif_menor
[1] 65
```

3.- Obtener la calificación promedio por unidad.

```
promedio_unidad <- apply(calificaciones, 2, mean)
promedio_unidad
```

```
> promedio_unidad
    U1     U2     U3     U4
84.125 83.750 81.125 82.625
>
```

4.- Obtener la calificación promedio por alumno.

```
# 4. Obtener la calificación promedio por alumno
promedio_alumno <- apply(calificaciones, 1, mean)
promedio_alumno
```

```
> promedio_alumno
 Fabian   Ramon    Luis Liliana   Diana   Erika  Karina   Jorge
  84.75   82.25   82.00   86.00   81.00   80.75   78.25   88.25
```

5.- Mostrar cuantas calificaciones están dadas de alta.

```
# 5. Mostrar cuantas calificaciones están dadas de alta
total_calificaciones <- length(calificaciones)
total_calificaciones
```

```
> total_calificaciones
[1] 32
```

6.- Obtener la mediana de las calificaciones.

```
# 6. Obtener la mediana de las calificaciones
mediana_calificaciones <- median(calificaciones)
mediana_calificaciones
```

```
> mediana_calificaciones
[1] 82
```

Practica ejercicios 3 unidad 3

Instrumento	***Práctica de ejercicios***

Alumno: Peña Ortiz Jose Alberto	**Fecha:** 25 de junio de 2024
Carrera: Ingeniería en Desarrollo y Gestión de Software	**Grupo:** IDGS91
Asignatura: Extracción de Conocimiento en Bases de Datos.	**Unidad temática:** III. Análisis supervisado
Profesor: MGTI. María Eugenia Guerrero Chan	

Contenido

I.- Ejercicios a resolver:

Instrucciones:

Realiza lo siguiente en el programa de RStudio.

1.- Calcula el área de un triángulo.

2.- Calcula el área de un rectángulo.

3.- Calcula el perímetro de un rectángulo.

Nota: Dejar evidencia en cada uno de los pasos con el código y con tus propias palabras.

II.-Procedimientos y resultados: (Poner aquí la estructura y orden de la información)

1.- Calcula el área de un triángulo.

Asignamos el valor 10 a la variable base_triangulo

```
base_triangulo <- 10
```

Asignamos el valor 7 a la variable altura_triangulo

```
altura_triangulo <- 7
```

Calculamos el área del triángulo utilizando la fórmula (base * altura) / 2

```
area_triangulo <- (base_triangulo * altura_triangulo) / 2
```

Mostramos el resultado del área del triángulo con los valores de base y altura

```
cat("El área del triángulo con base", base_triangulo, "y altura", altura_triangulo, "es:", area_triangulo, "\n\n")
```

Código completo del área de un triangulo

```
# 1.- Calcula el área de un triángulo
# Fórmula: Área = (base * altura) / 2
cat("Cálculo del área de un triángulo\n")
base_triangulo <- 10
altura_triangulo <- 7
area_triangulo <- (base_triangulo * altura_triangulo) / 2
cat("El área del triángulo con base", base_triangulo, "y altura", altura_triangulo, "es:", area_triangulo, "\n\n")
```

Consola del área de un triangulo

```
> cat("Cálculo del área de un triángulo\n")
Cálculo del área de un triángulo
> base_triangulo <- 10
> altura_triangulo <- 7
> area_triangulo <- (base_triangulo * altura_triangulo) / 2
> cat("El área del triángulo con base", base_triangulo, "y altura", altura_triangulo, "es:", area_triangulo, "\n\n")
El área del triángulo con base 10 y altura 7 es: 35
```

2.- Calcula el área de un rectángulo.

Fórmula: Área = base * altura

Se definen las variables base_rectangulo y altura_rectangulo con valores de 15 y 8 respectivamente.

```
base_rectangulo <- 15

altura_rectangulo <- 8
```

Se calcula el área del rectángulo usando la fórmula base * altura y se guarda en la variable area_rectangulo.

```
area_rectangulo <- base_rectangulo * altura_rectangulo
```

Se muestra el resultado usando cat(), indicando los valores de base, altura y el área calculada.

```
cat("El área del rectángulo con base", base_rectangulo, "y altura", altura_rectangulo, "es:", area_rectangulo, "\n\n")
```

Código completo del área de un rectángulo

```
# 2.- Calcula el área de un rectángulo
# Fórmula: Área = base * altura
cat("Cálculo del área de un rectángulo\n")
base_rectangulo <- 15
altura_rectangulo <- 8
area_rectangulo <- base_rectangulo * altura_rectangulo
cat("El área del rectángulo con base", base_rectangulo, "y altura", altura_rectangulo, "es:", area_rectangulo, "\n\n")
```

Consola del área de un rectángulo

```
Cálculo del área de un rectángulo
> base_rectangulo <- 15
> altura_rectangulo <- 8
> area_rectangulo <- base_rectangulo * altura_rectangulo
> cat("El área del rectángulo con base", base_rectangulo, "y altura", altura_rectangulo, "es:", area_rectangulo, "\n\n")
El área del rectángulo con base 15 y altura 8 es: 120
```

3.- Calcula el perímetro de un rectángulo.

Fórmula: Perímetro = 2 * (base + altura)

Se calcula el perímetro usando la fórmula 2 * (base + altura) y se guarda en la variable perimetro_rectangulo.

```
perimetro_rectangulo <- 2 * (base_rectangulo + altura_rectangulo)
```

Se muestra el resultado usando cat(), indicando los valores de base, altura y el perímetro calculado.

```
cat("El perímetro del rectángulo con base", base_rectangulo, "y altura", altura_rectangulo, "es:", perimetro_rectangulo,
```

Código completo del perímetro de un rectángulo

```
# 3.- Calcula el perímetro de un rectángulo
# Fórmula: Perímetro = 2 * (base + altura)
cat("Cálculo del perímetro de un rectángulo\n")
perimetro_rectangulo <- 2 * (base_rectangulo + altura_rectangulo)
cat("El perímetro del rectángulo con base", base_rectangulo, "y altura", altura_rectangulo, "es:", perimetro_rectangulo,
```

Consola del perímetro de un rectángulo

```
> cat("Cálculo del perímetro de un rectángulo\n")
Cálculo del perímetro de un rectángulo
> perimetro_rectangulo <- 2 * (base_rectangulo + altura_rectangulo)
> cat("El perímetro del rectángulo con base", base_rectangulo, "y altura", altura_rectangulo, "es:", perimetro_rectangulo, "\n
\n")
El perímetro del rectángulo con base 15 y altura 8 es: 46
```

Examen Unidad 3

Universidad Tecnológica de la Costa
Gobierno del Estado de Nayarit
Organismo Público Descentralizado

Examen
Sistema de Gestión de la Calidad

Fecha de emisión: 10/01/2013
Revisión: 01
Página **1** de **2**

Instrumento	Examen

Alumno: Peña ortiz Jose Alberto	**Fecha:** 09107124
Carrera: Ingeniería en Desarrollo y Gestión de Software	**Grupo:** IDGS91
Asignatura: Extracción de Conocimiento en Bases de Datos.	**Unidad temática:** III. Análisis supervisado
Profesor: MGTI. María Eugenia Guerrero Chan	

I. Reactivos

70

1. **Es una rama de machine learning, método de análisis de datos que trabaja con datos etiquetados con base a un histórico crea un modelo que realiza predicciones basadas en evidencia en presencia de incertidumbre.**

a) Análisis no supervisado
b) Análisis semisupervisado
c) Análisis supervisado
d) Ninguno de los anteriores

2. **Estos algoritmos, basados en técnicas de aprendizaje supervisado, permiten clasificar o etiquetar datos en clases predefinidas, utilizando las características específicas de cada elemento para determinar su pertenencia a una clase particular.**

a) Algoritmo de regresión
b) Algoritmo de clasificación
c) Algoritmo de transformación
d) Algoritmo de reducción

3. **Algoritmo de aprendizaje supervisado que se utiliza en ML y en estadística y, en términos sencillos, establece una recta para proporcionar la tendencia de un conjunto de datos.**

a) Algoritmo de clasificación
b) Algoritmo de agrupación
c) Algoritmo de regresión
d) Algoritmo de reducción

X

4. **Es el ejemplo más común de algoritmos de clasificación:**

a) Redes sociales
b) El detector de correo no deseado del correo electrónico
c) Data warehouse
d) Datos estructurados

5. **Son ejemplos de los usos de los algoritmos de regresión, excepto:**

a) Predicción del precio de la vivienda
b) Correo electrónico
c) El tiempo de permanencia de un empleado en una empresa
d) Estimación de ventas de productos

6. **Se fundamenta en la identificación de relaciones entre variables en eventos pasados, para luego explotar dichas relaciones y predecir posibles resultados en futuras situaciones.**

a) Análisis inferencial
b) Análisis descriptivo
c) Análisis predictivo
d) Ninguna de las anteriores

7. **¿Cuál es el orden de los pasos básicos del aprendizaje supervisado?**

1- Elegir un modelo apropiado, 2- Ajustar y afinar el modelo, 3- Entrenar el modelo, 4- Evaluar el modelo, 5- Preparación y preprocesamiento de datos, 6- Recopilar datos etiquetados, 7- Hacer predicciones sobre nuevos datos.

a) 6,5,1,3,4,2,7 b) 2,4,3,1,6,5,7 c) 1,2,3,4,5,6,7 d) 6,5,1,2,3,4,7

8. **Son ejemplos de algoritmos de clasificación, excepto:**

a) KNN b) Naive Bayes c) Arboles de decisión d) Diagrama de flujo

9. **Es un algoritmo que modela la relación entre una variable dependiente y una o más variables independientes, se utiliza en la predicción de precios y tendencias.**

a) Regresión lineal b) Regresión logística c) Big Data d) Ninguna de las anteriores

10. **Son herramientas que permiten medir y cuantificar el rendimiento de un modelo**.

a) Métricas de evaluación b) Machine learning c) KNN d) Big data

Conclusión