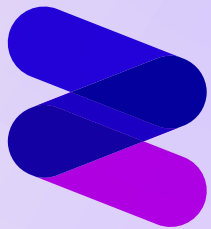


cedia



research  
lab

by cedia

**Redacción y diseño de  
los casos de uso de las  
herramientas HPC, Indata,  
GitHub y Mattermost**

—  
CASO DE  
USO 2

cedia

*“Conectando ideas,  
transformando  
sociedades”*





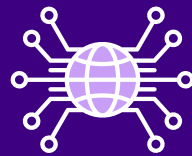
Almacena y comparte  
datos de investigación.



Usa computación  
de alto rendimiento.



Colabora en proyectos de  
investigación de manera  
efectiva.



Envía archivos  
grandes de forma  
segura.

*Mejorará la  
reproducibilidad  
y eficiencia en  
tu investigación*



*Una Plataforma que revolucionará  
la Investigación Científica en Ecuador*



**research  
lab**  
by cedia

Redacción y diseño de los casos  
de uso de las herramientas HPC,  
Indata, GitHub y Mattermost

CASO DE USO 2

# CONTENIDO

## CASO DE USO 2

1.1

Objetivo

pp – 08

1.2

Descripción

pp – 08

1.3



Uso de la herramienta Indata

pp – 20

1.4



Uso de la herramienta HPC

pp – 10

1.5



Uso de la herramienta Gitlab

pp – 46

1.6



Uso de la herramienta Mattermost

pp – 60

1.7



Uso de la herramienta Filesender

pp – 72

# CASO DE USO 2

## Análisis Geoespacial y Temporal de Transacciones

### 1.1

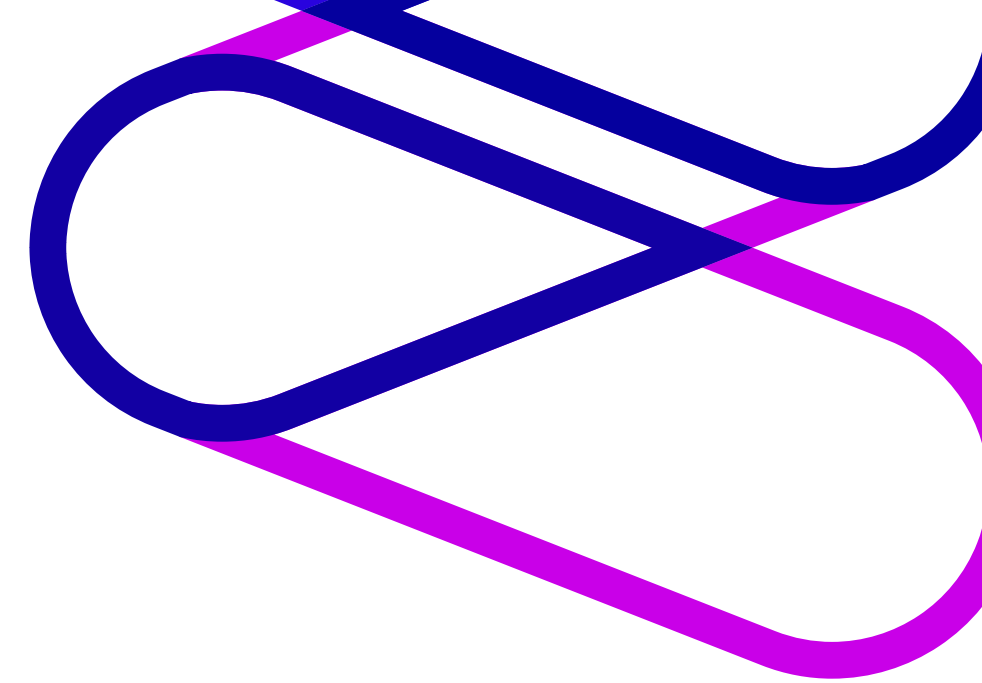
#### OBJETIVO

Realizar un análisis geoespacial y temporal de las transacciones para detectar patrones de comportamiento del cliente, posibles fraudes, y tendencias en diferentes ubicaciones y horarios, utilizando el clúster **HPC** para el procesamiento paralelo de los datos.

### 1.2

#### DESCRIPCIÓN

Este proyecto se enfoca en analizar el dataset de transacciones utilizando datos geoespaciales (latitud y longitud) y temporales (fecha y hora de la transacción). Se pueden explorar patrones en diferentes ubicaciones geográficas, horarios pico de transacciones, tasas de fraude según la ubicación y la hora del día, entre otros aspectos. Se utilizará el clúster HPC para realizar análisis a gran escala y GitHub para el control de versiones y colaboración. Además, InData se puede usar para visualización y almacenamiento, y Mattermost para la comunicación y coordinación del equipo.



# 1.3

## USO DE LA HERRAMIENTA INDATA



### Dataset

**Nombre:** ShopSpectra Transaction Dataset.csv

**Descripción:** Dataset con múltiples transacciones que incluye columnas sobre montos, ubicación geográfica, dispositivos, métodos de pago, etc.

**Tamaño:** 145.72 MB

**Referencia:** Mishra, A. (2023). ShopSpectra - E-commerce Transactions [Data set]. Kaggle. <https://www.kaggle.com/datasets/mishraarushi02/shopspectra-e-commerce-transactions>

### Adquisición de datos

En este caso, utilizaremos un dataset que contiene características:

TransactionID: Unique identifier for each transaction.

UserID: Unique identifier for each customer.

TransactionAmount: The monetary value of the transaction.

TransactionDate1: (Please ignore/delete this column).

PaymentMethod: Method used for payment (e.g., PayPal, Credit Card, Debit Card).

MerchantCategory: Category of the purchased items (e.g., Electronics, Clothing, Entertainment).

Quantity: Number of items purchased in the transaction.

CustomerAge: Age of the customer.

Location: City where the transaction occurred.

DeviceType: Type of device used for the transaction (e.g., Mobile, Desktop, Tablet).

TransactionStatus: Current status of the transaction (e.g., Completed, Pending, Failed).

Is\_Declined: Binary indicator (0 or 1) if the transaction was declined.

Is\_Fraud: Binary indicator (0 or 1) if the transaction was fraudulent.

AccountAgeDays: Age of the customer's account in days.

TransactionDate: Date of the transaction in MM/DD/YY format.

Latitude: Latitude coordinate of the transaction location.

Longitude: Longitude coordinate of the transaction location.

TransactionHour: Hour of the day when the transaction occurred (0-23).



TransactionID	UserID	TransactionAmount	TransactionDate	PaymentMethod	MerchantCategory	Quantity	CustomerAge	Location	DeviceType	TransactionStatus	Is_Declined	Is_Fraud	AccountAgeDays	TransactionDate	Latitude	Longitude	TransactionHour
1,29318,1307.37	2023-12-19	19:47:58	775599	PayPal	Entertainment	5	79	Oklahoma City	Tablet	Pending	0	0	31	2023-06-10	35.4676	-97.5194	8
2,33379,11643.05	2023-09-15	08:43:58	775599	Debit Card	Electronics	8	7	Chicago	Mobile	Completed	0	0	128	2023-09-16	41.8781	-87.6298	19
3,25503,2007.01	2023-07-26	13:48:58	775599	PayPal	Electronics	1	70	Nashville	Desktop	Completed	0	0	15	2023-04-14	36.1627	-86.7816	3
4,7330,4796.81	2023-12-30	18:32:58	775599	Credit Card	Home & Garden	5	18	Chicago	Desktop	Completed	0	0	324	2023-05-30	41.8781	-87.6298	19
5,25851,5992.9	2024-03-19	00:00:58	775599	Credit Card	Clothing	8	12	Austin	Mobile	Pending	0	0	311	2023-12-08	30.2672	-97.7431	19
6,987,8214.46	2023-12-03	14:26:58	775599	Credit Card	Clothing	7	11	Dallas	Desktop	Completed	0	0	262	2024-04-09	32.7767	-96.7974	4
7,27437,1221.82	2023-11-30	03:52:58	775599	Debit Card	Electronics	10	4	San Jose	Mobile	Completed	0	1	353	2023-05-11	37.3382	-121.8863	13
8,3770,1864.98	2024-03-26	03:21:58	775599	Debit Card	Entertainment	10	61	Denver	Mobile	Completed	0	0	2	2023-11-05	39.7392	-104.9903	10
9,12261,11348.14	2024-05-03	13:48:58	775599	PayPal	Entertainment	10	53	New York	Mobile	Pending	0	0	112	2023-11-19	40.7128	-74.006	14
10,27457,206.76	2023-12-26	06:35:58	775599	Debit Card	Entertainment	6	84	Fort Worth	Mobile	Completed	0	0	152	2023-04-16	32.7555	-97.3308	5
11,11488,8986.55	2024-01-20	16:44:58	775599	Credit Card	Travel	1	6	San Jose	Mobile	Pending	0	0	260	2024-01-25	37.3382	-121.8863	16
12,6855,2240.71	2023-09-13	23:00:58	775599	PayPal	Travel	6	50	New York	Desktop	Completed	0	0	148	2023-11-07	40.7128	-74.0064	4
13,21703,6442.73	2023-11-14	13:14:58	775599	PayPal	Electronics	4	15	Los Angeles	Tablet	Completed	0	0	209	2023-01-29	34.0522	-118.2437	8
14,145648,3623.16	2023-10-22	08:27:58	775599	Debit Card	Entertainment	9	66	Charlotte	Tablet	Completed	0	0	301	2023-11-01	35.2271	-80.8431	16
15,17786,13812.52	2023-10-15	23:54:58	775599	Credit Card	Entertainment	8	12	Nashville	Mobile	Completed	0	1	202	2023-04-22	36.1627	-86.7816	11
16,26261,11479.09	2024-02-03	23:23:58	775599	Debit Card	Groceries	5	79	Charlotte	Mobile	Completed	0	0	58	2023-07-21	35.2271	-80.8431	8
17,3275,110055.3	2023-12-29	16:59:58	775599	Credit Card	Entertainment	8	82	Columbus	Mobile	Failed	0	0	345	2023-12-15	39.9612	-82.9988	6
18,21629,10835.26	2023-12-30	18:34:58	775599	Credit Card	Clothing	4	7	Los Angeles	Tablet	Pending	0	0	130	2024-05-18	34.0522	-118.2437	7
19,26277,7508.61	2023-11-24	12:16:58	775599	Bank Transfer	Groceries	4	24	Phoenix	Mobile	Completed	0	0	231	2023-03-05	33.4484	-112.074	13
20,8847,1082.11	2024-04-20	05:56:58	775599	Bank Transfer	Clothing	7	82	San Antonio	Tablet	Pending	0	0	227	2023-11-09	29.4241	-98.4936	4
21,33012,14583.02	2024-03-08	00:39:58	775599	Credit Card	Toys & Games	4	63	Phoenix	Desktop	Completed	0	0	60	2024-02-23	33.4484	-112.074	16
22,33998,7268.32	2024-01-11	14:47:58	775599	Credit Card	Groceries	3	57	San Francisco	Desktop	Completed	0	0	157	2023-06-17	37.7749	-122.4194	0
23,3706,5990.07	2024-05-06	18:36:58	775599	Credit Card	Home & Garden	7	62	Columbus	Tablet	Completed	0	0	2	2023-07-02	39.9612	-82.9988	12
24,26943,11295.21	2023-06-30	01:55:58	775599	PayPal	Electronics	6	39	San Jose	Mobile	Completed	0	0	270	2024-04-16	37.3382	-121.8863	15
25,32790,14256.95	2024-05-19	05:29:58	775599	Credit Card	Entertainment	1	45	Detroit	Desktop	Pending	1	0	228	2024-03-24	42.3314	-83.0458	11
26,32398,12109.45	2023-06-24	00:14:58	775599	Credit Card	Clothing	6	27	El Paso	Mobile	Failed	0	0	278	2023-11-19	31.7619	-106.485	0
27,37000,13634.36	2023-06-24	00:14:58	775599	Credit Card	Clothing	6	27	El Paso	Mobile	Failed	0	0	278	2023-11-19	31.7619	-106.485	0

figura 1

Incluimos los resultados dentro del dataset en la aplicación InData, permitiendo que otros investigadores accedan a ellos para validar y comprobar la eficiencia del modelo de entrenamiento de aprendizaje automático. Asimismo, añadimos la metadata relacionada con la Geospatial Metadata, ya que el dataset contiene información geoespacial (latitud y longitud) y temporal (fecha y hora de la transacción).

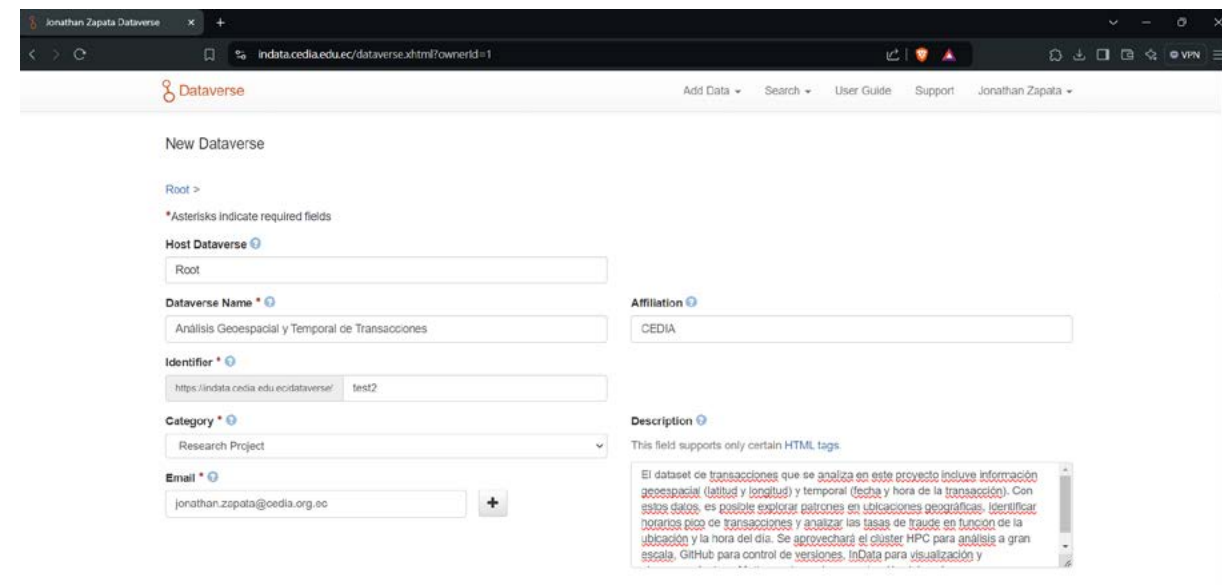


figura 2

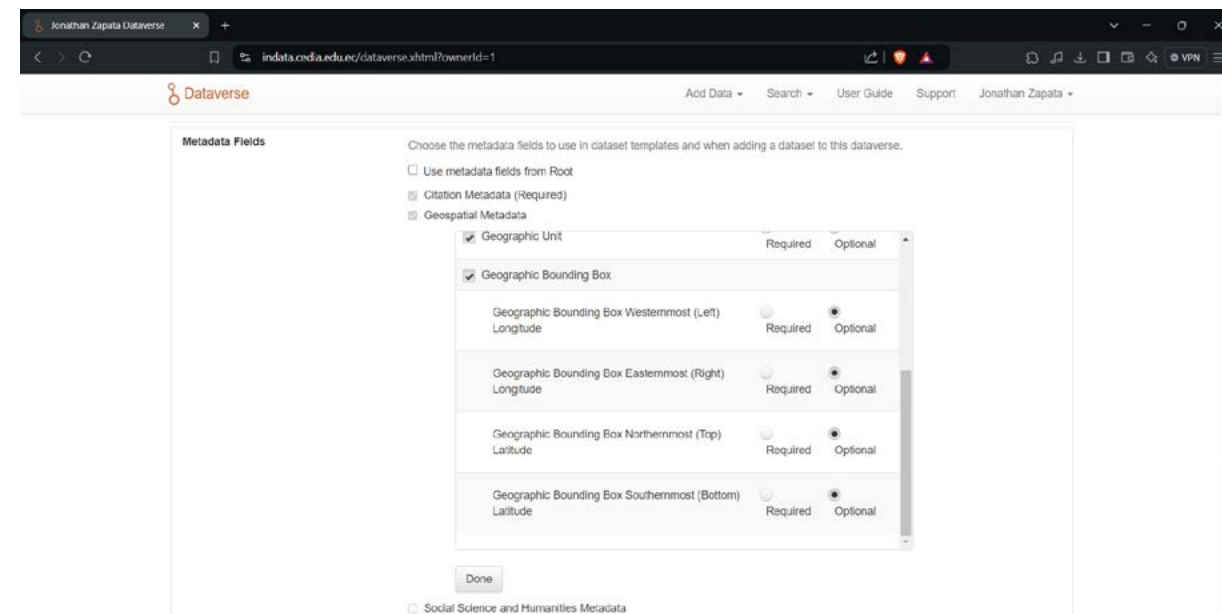


figura 3



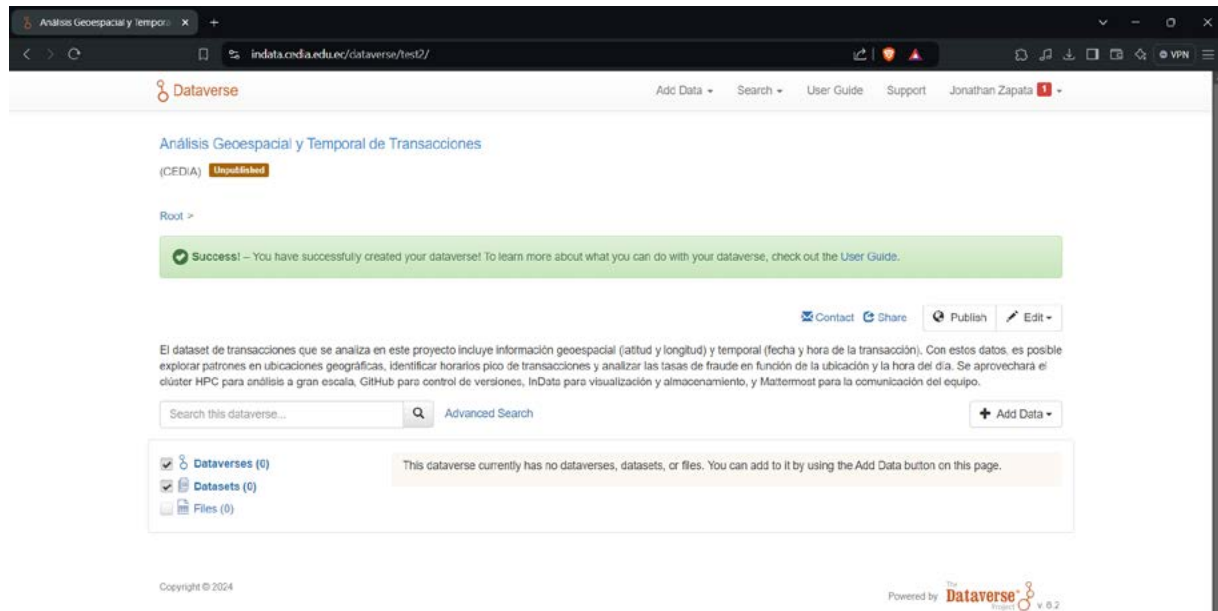


figura 4

Editamos los accesos, para que mas investigadores puedan acceder a nuestro dataset

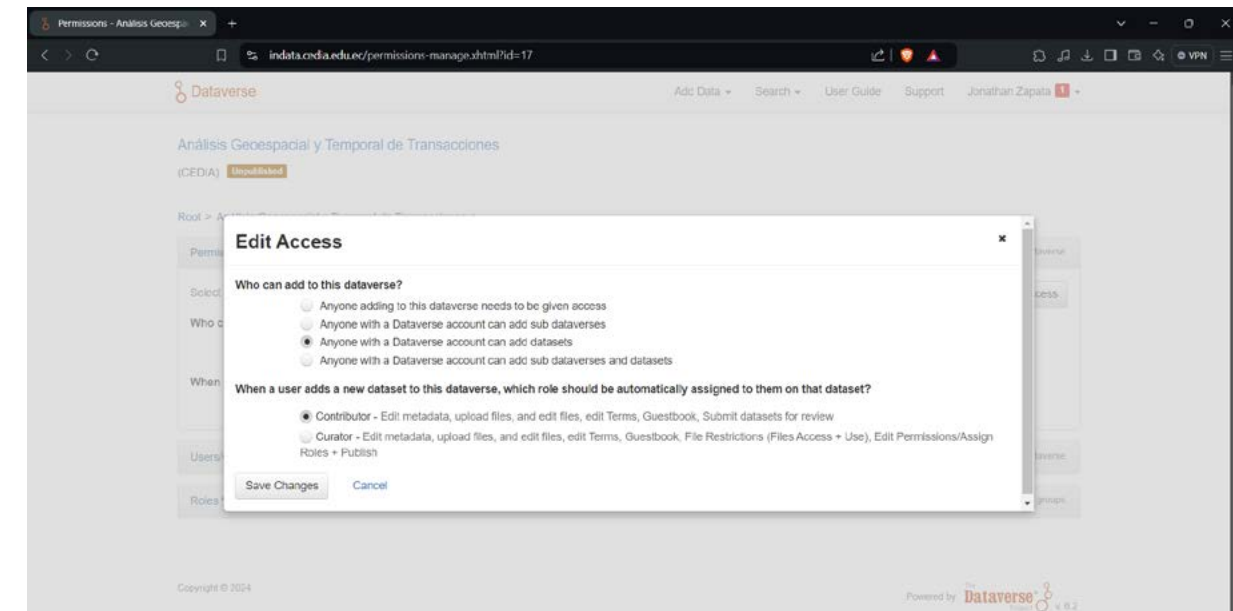


figura 6

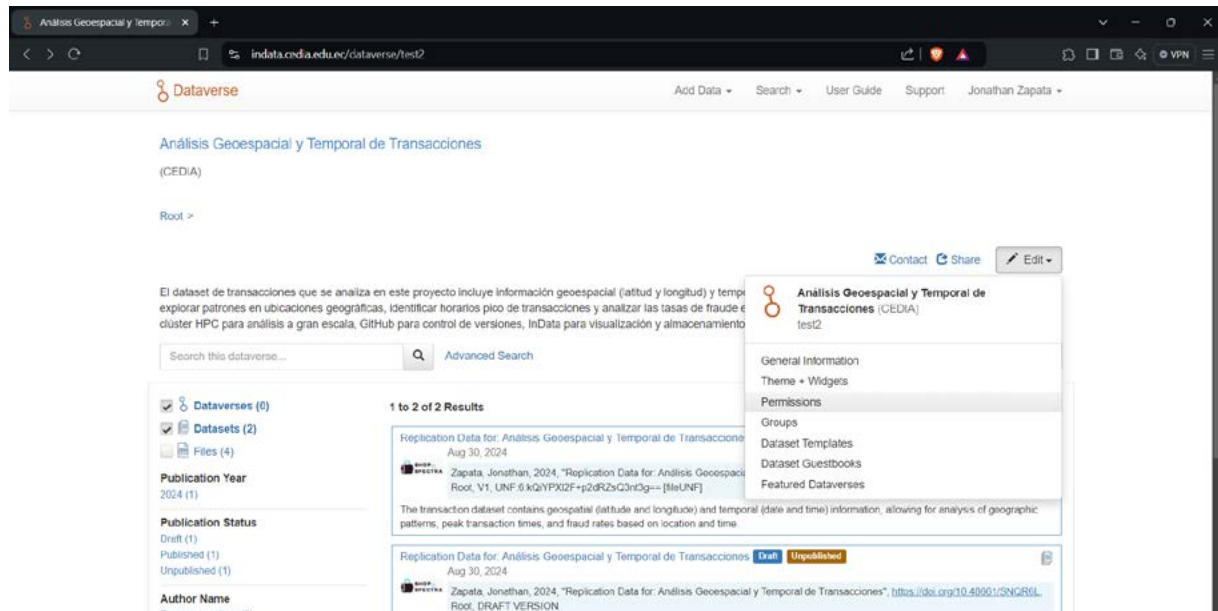


figura 5

Y finalmente publicamos nuestro repositorio de datos en dataverse

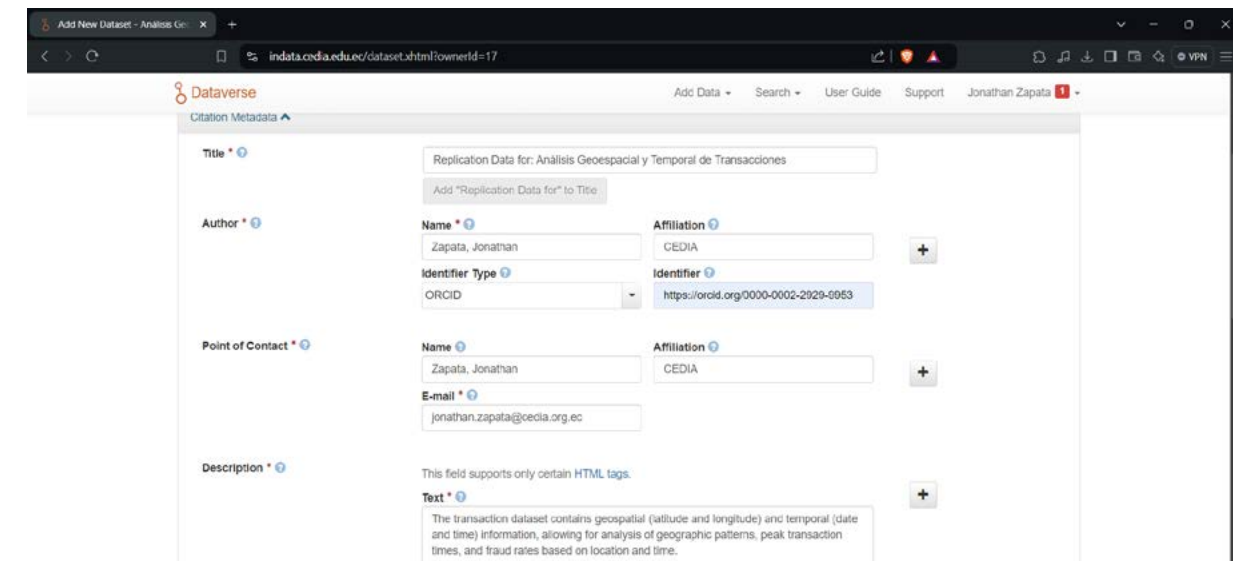


figura 7

The screenshot shows the 'Add New Dataset' form in Dataverse. The 'Description' field contains the text: 'The transaction dataset contains geospatial (altitude and longitude) and temporal (date and time) information, allowing for analysis of geographic patterns, peak transaction times, and fraud rates based on location and time.' The 'Date' field is set to '2024-08-30'. The 'Subject' dropdown is set to 'Business and Management'. The 'Keyword' section includes two entries: 'Transaction' with 'Financial Industry Business Ontology (FI)' and 'Payment Method' with 'Financial Industry Business Ontology (FI)'. Both entries have their respective 'Controlled Vocabulary URL' set to 'https://www.edmouncil.org/fibo'.

figura 8

En este caso el siguiente dataset no fue utilizado por ningún artículo caso contrario al Caso de Uso 1, por ende los espacios se dejan en blanco.

The screenshot shows the 'Add New Dataset' form in Dataverse. The 'Related Publication' section includes a 'Citation' field, an 'Identifier Type' dropdown set to 'Select...', and an 'Identifier' field. The 'URL' field is set to 'https://'. The 'Notes' field is empty. The 'Depositor' field is set to 'Zapata, Jonathan' and the 'Deposit Date' field is set to '2024-08-30'.

figura 9





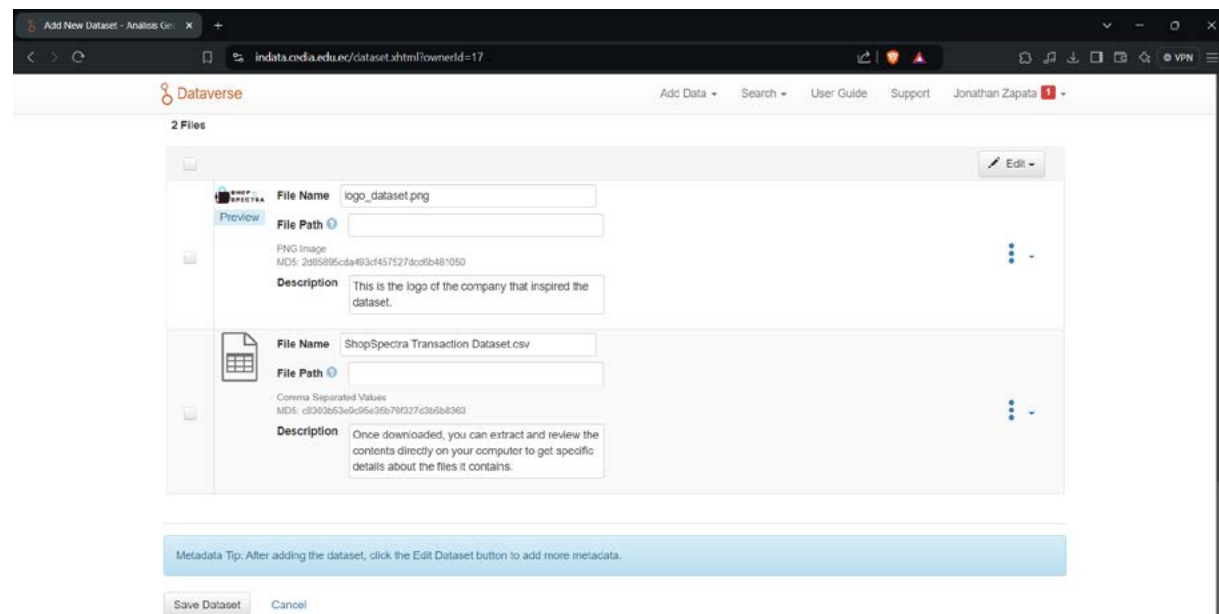


figura 10

Y finalmente creado el dataset dentro de nuestro repositorio o mejor dicho en el Dataverse.

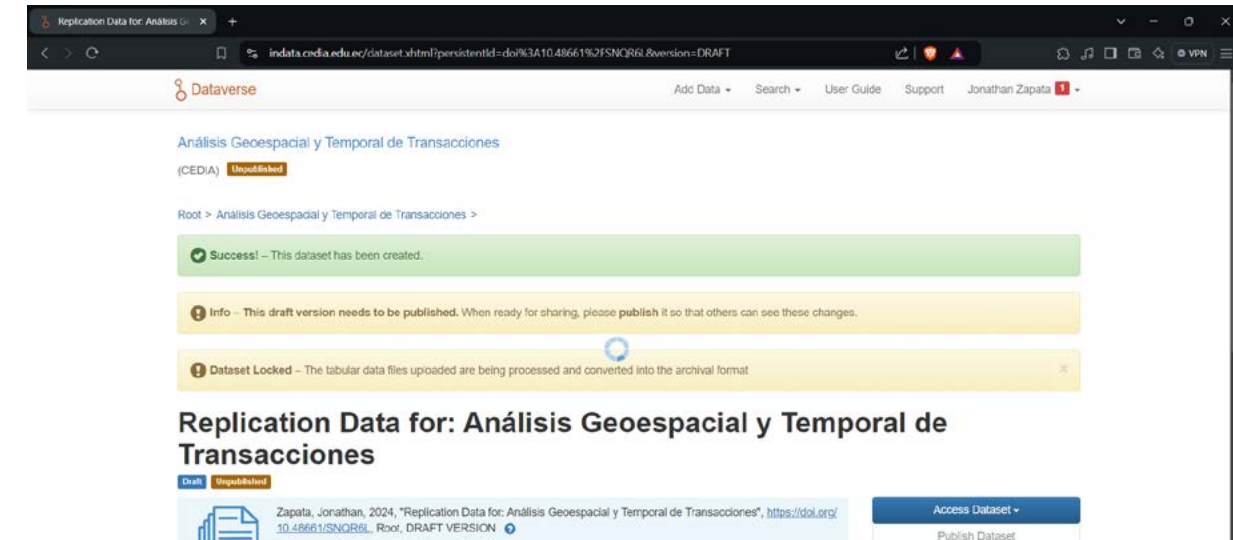


figura 11

Para que nuestro dataverse puedan verlo otros investigadores que buscan mediante un repositorio de Indata CEDIA lo publicamos para que sea publico y de libre acceso.

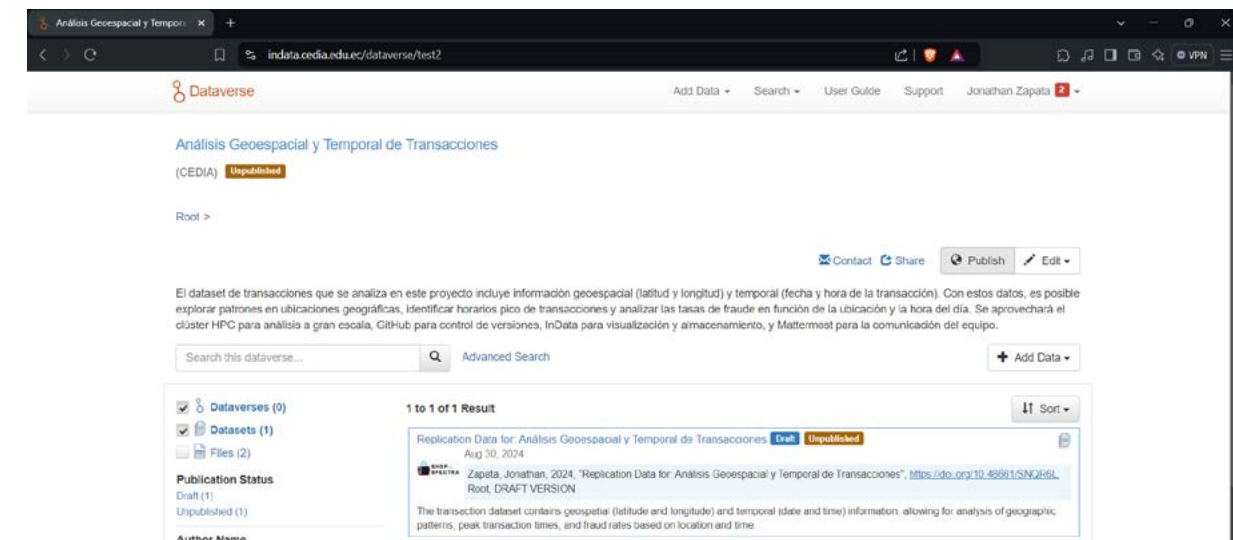


figura 12

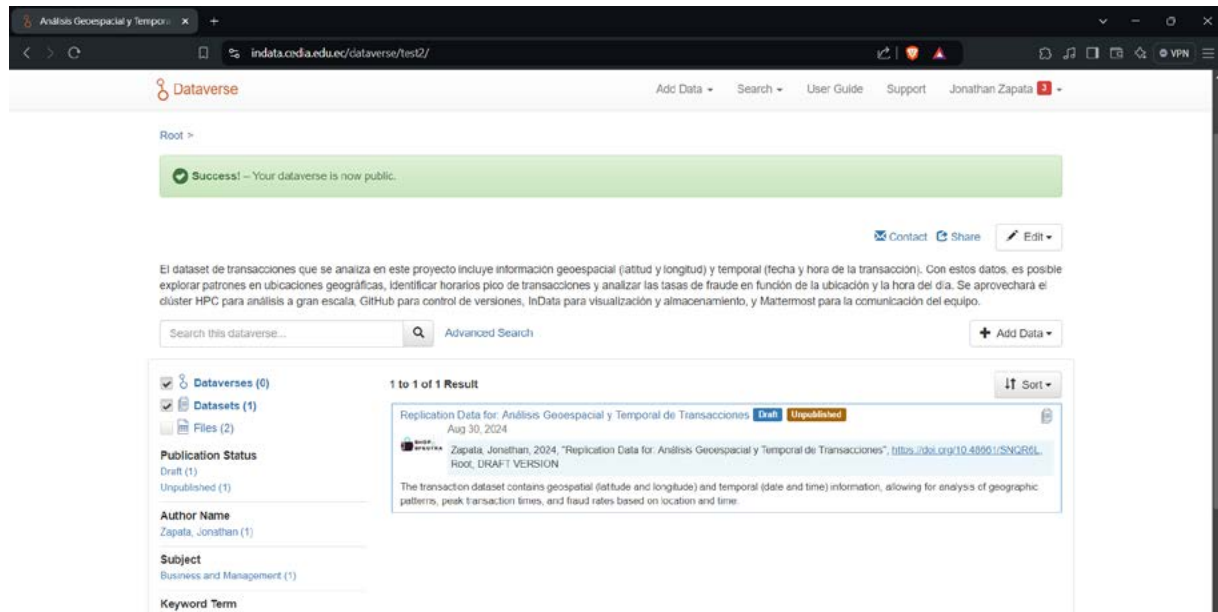


figura 13

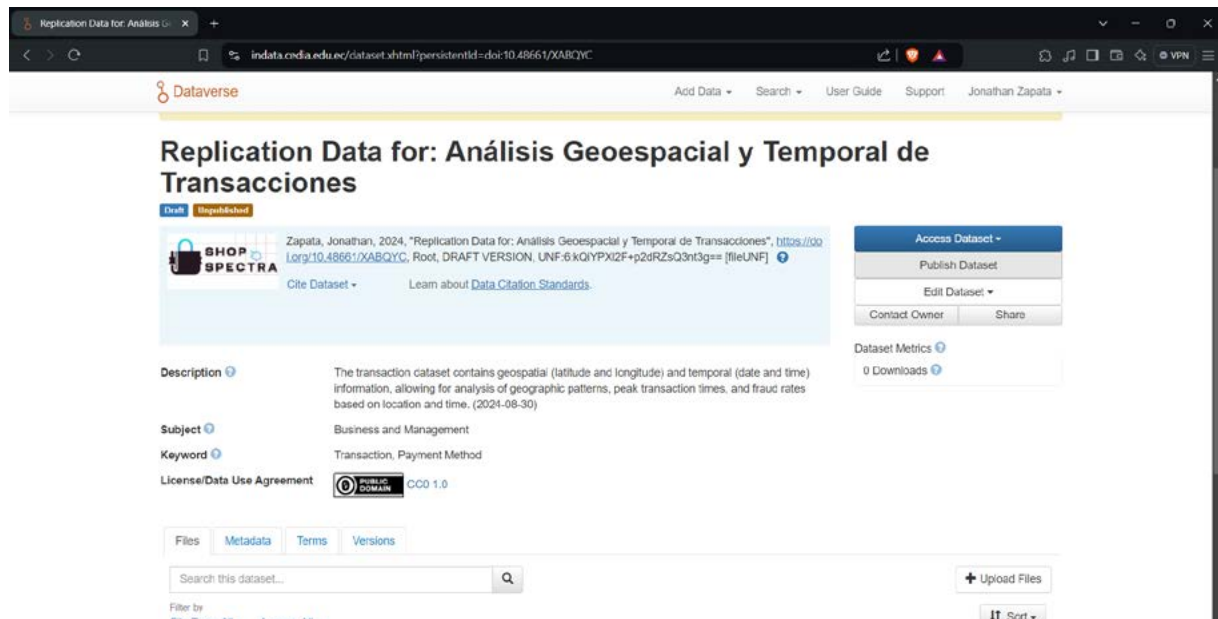


figura 14





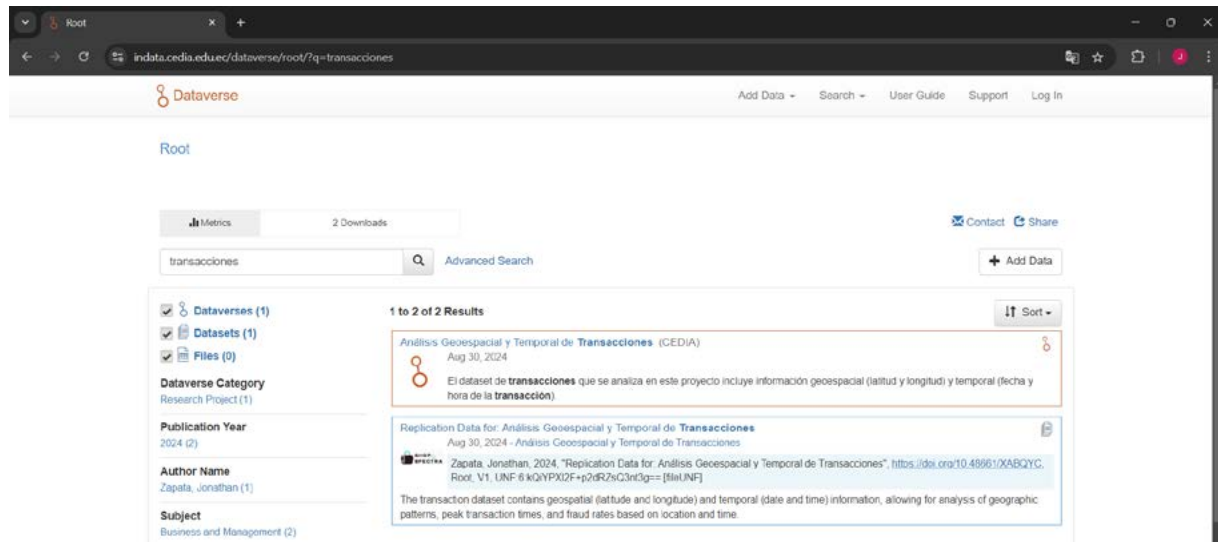


figura 15

Una vez que el archivo ha sido subido, podemos publicar nuestro dataset y compartirlo a través de diversos canales de comunicación, como Facebook, Twitter y LinkedIn.

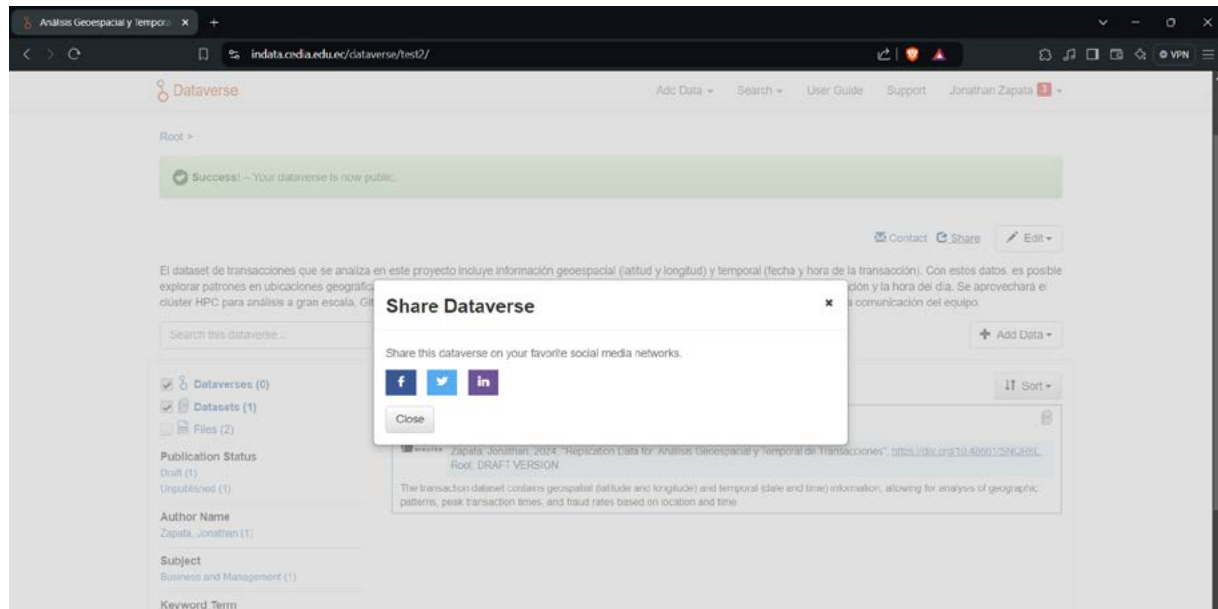


figura 16

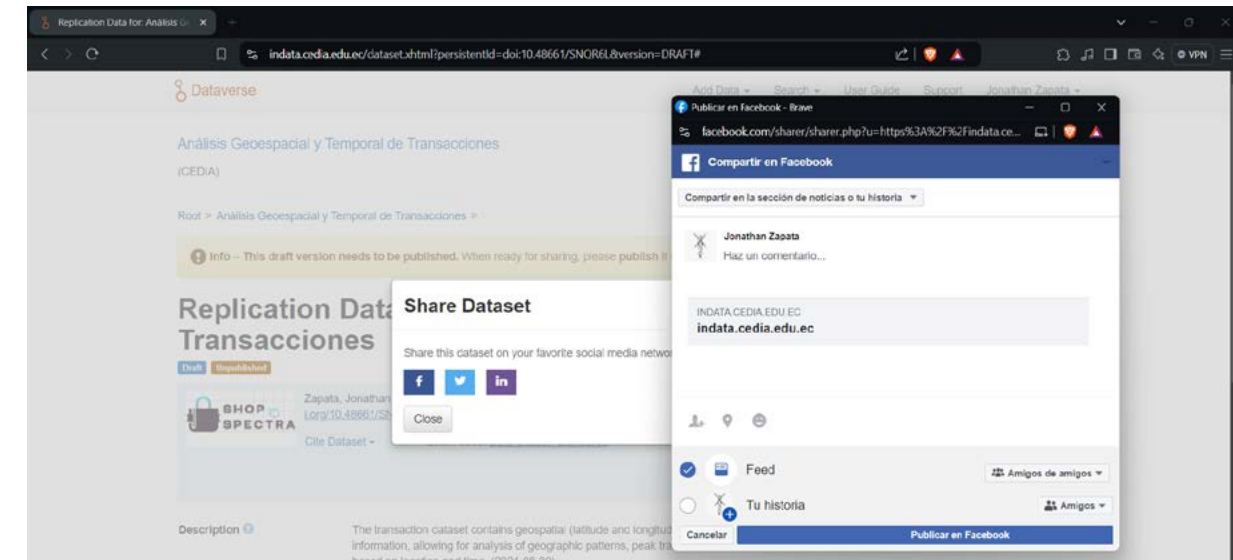


figura 17

Del mismo tenemos la opción de editar, eliminar y agregar el dataset del mismo modo los metadatos que acompañan a este:

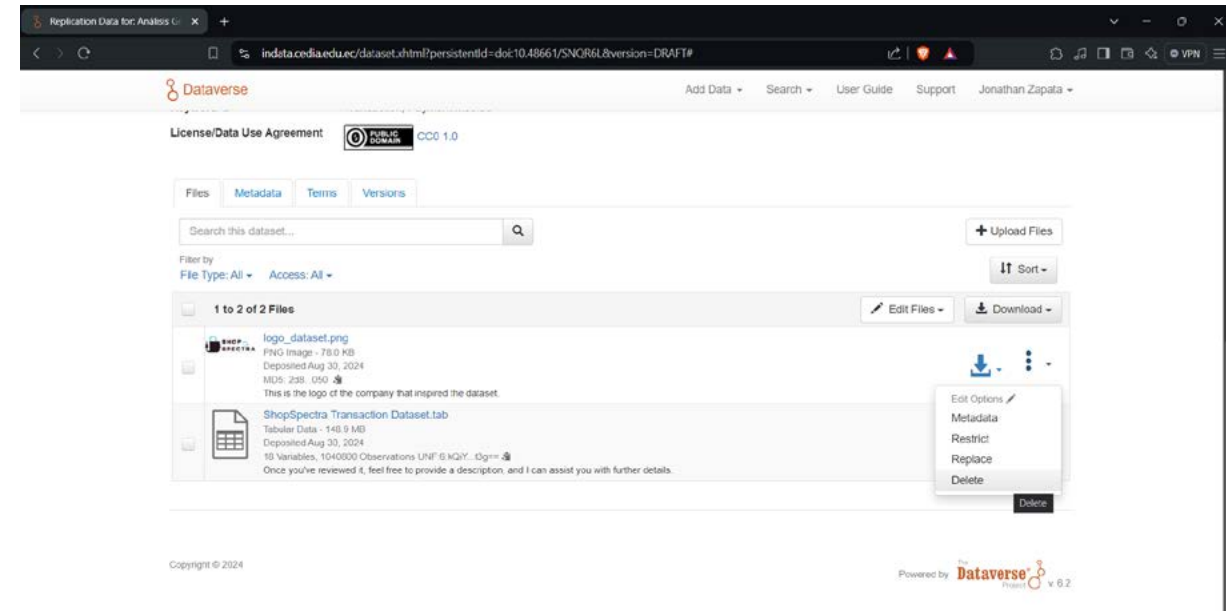


figura 18

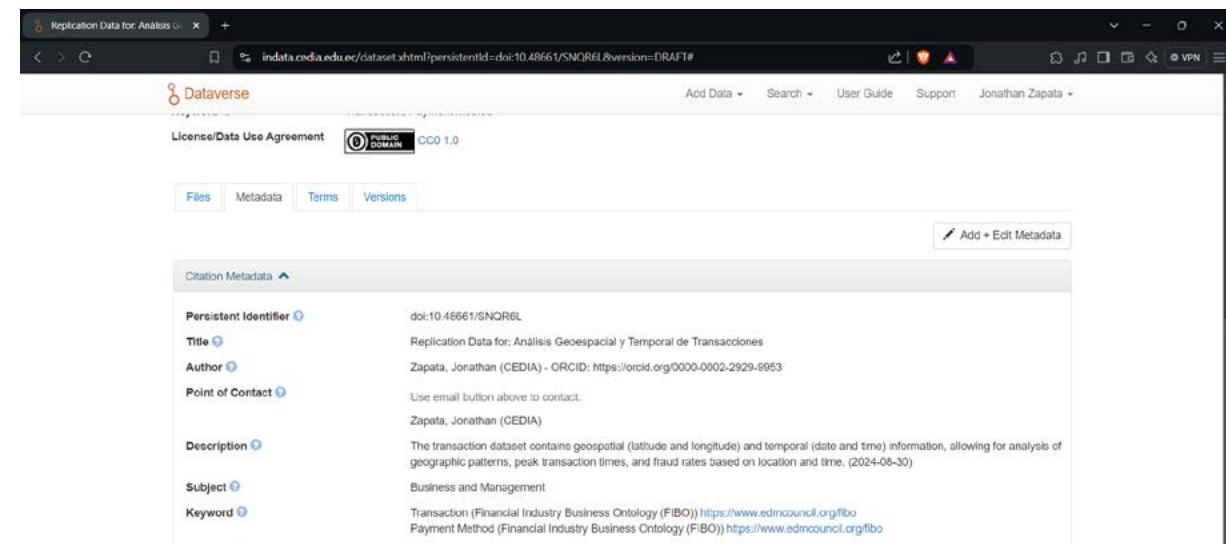


figura 19

Finalmente, nuestro Dataverse es público y accesible dentro de InDATA CEDIA, sin necesidad de realizar un login.



figura 20

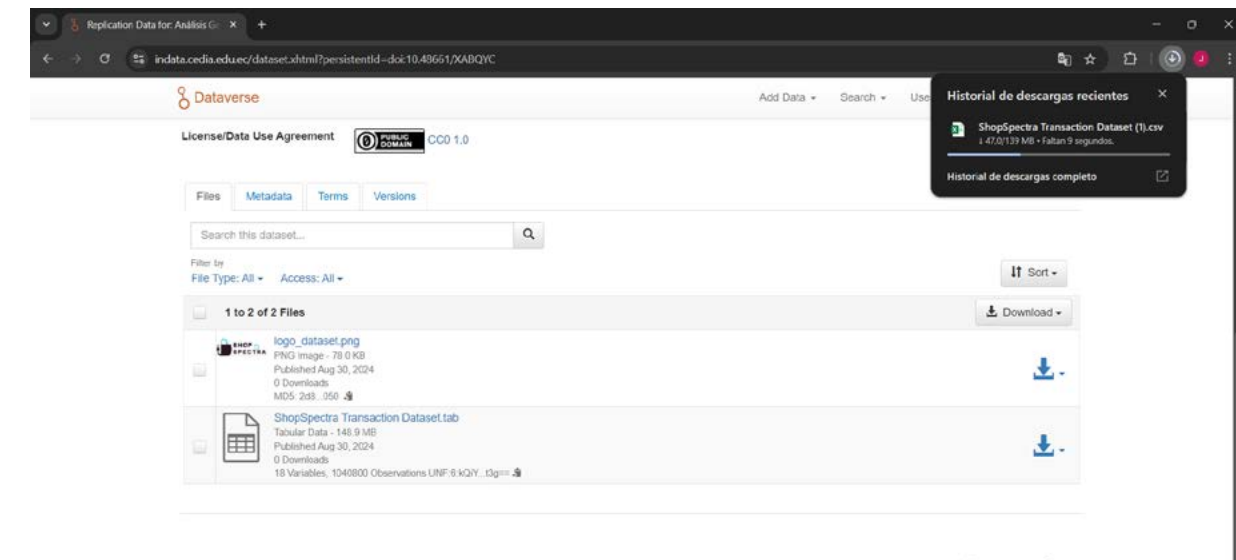


figura 24



# 1.3

## USO DE LA HERRAMIENTA HPC



# HPC

Creación de nuestro clúster para realizar procedimientos de análisis geoespacial y temporal básico con el dataset de transacciones. Este código se enfocará en generar estadísticas descriptivas, analizar las transacciones por hora y llevar a cabo un análisis geoespacial simple con gráficos. Y creamos el clúster de Jupyter con las siguientes recursos

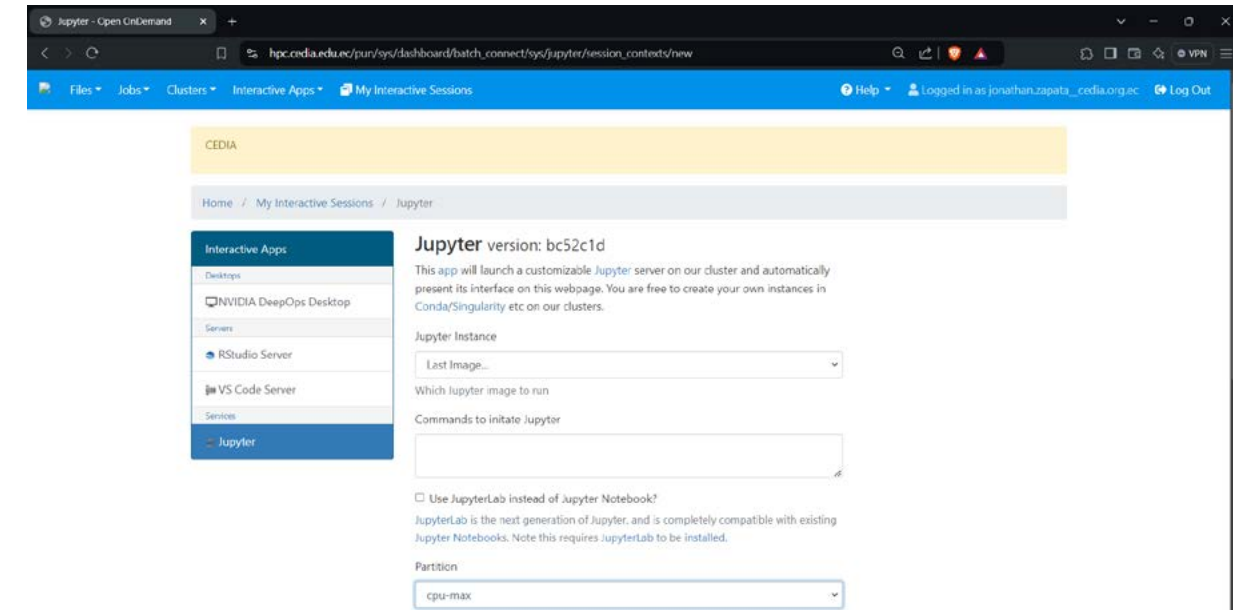


figura 25

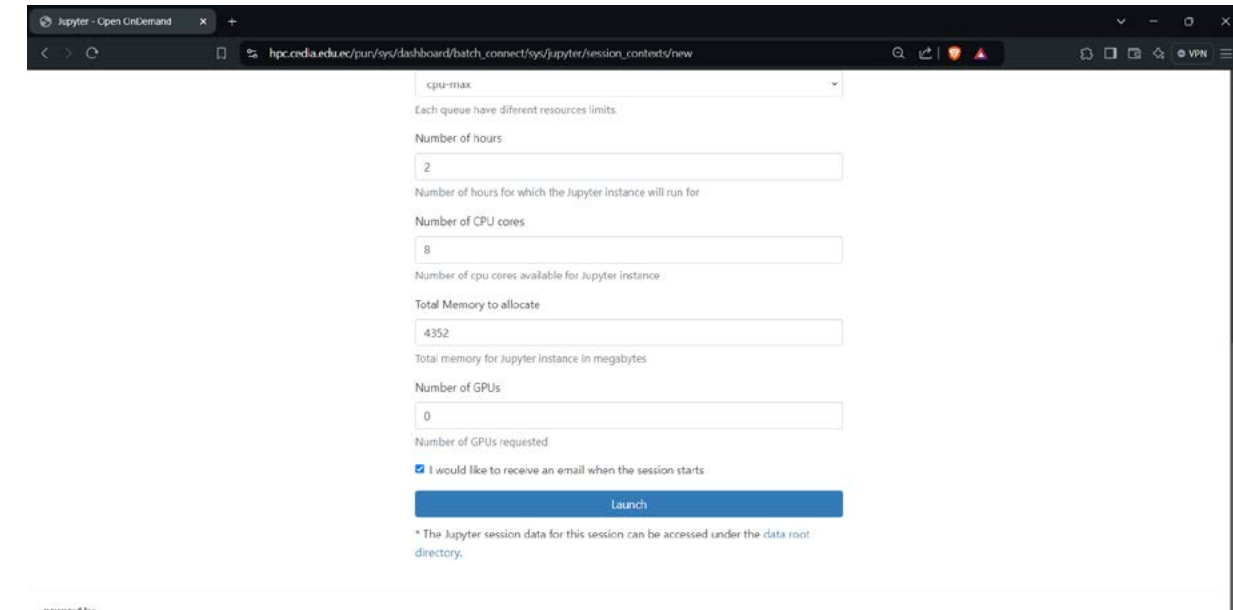


figura 26

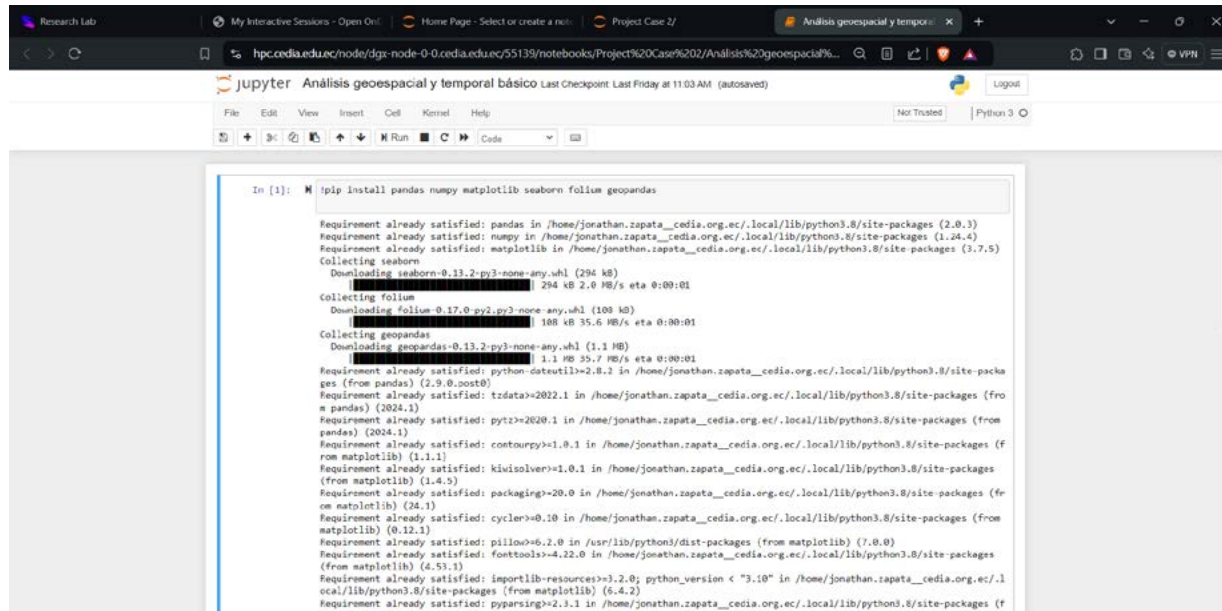


figura 27

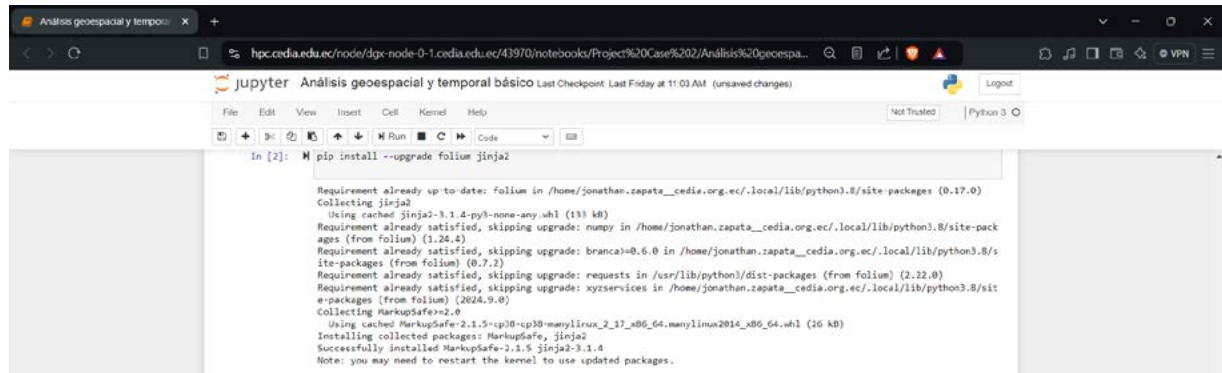


figura 28

Subir el dataset a nuestro entorno de trabajo en Jupyter.

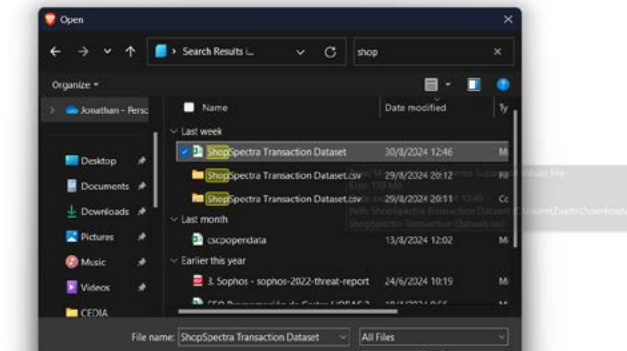


figura 29

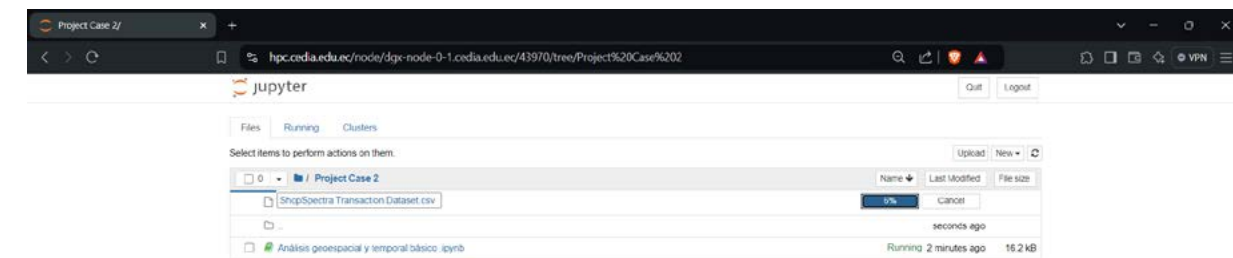


figura 30

Ejecución del código, junto con los estudios, análisis e investigaciones planificados.

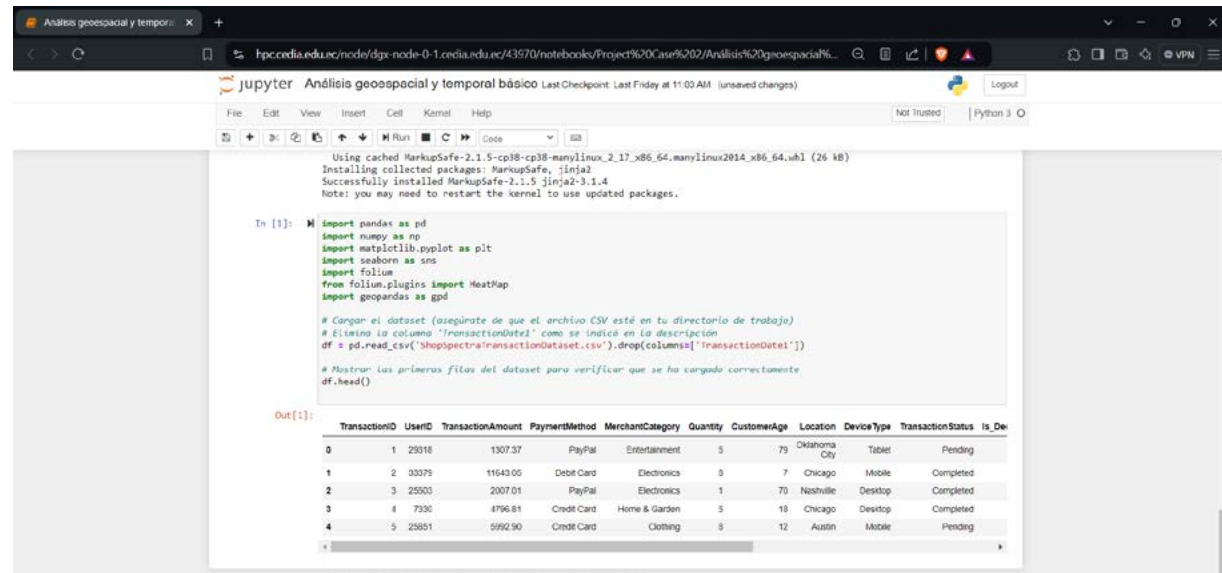


figura 31

## 1. Estadísticas Descriptivas Básicas

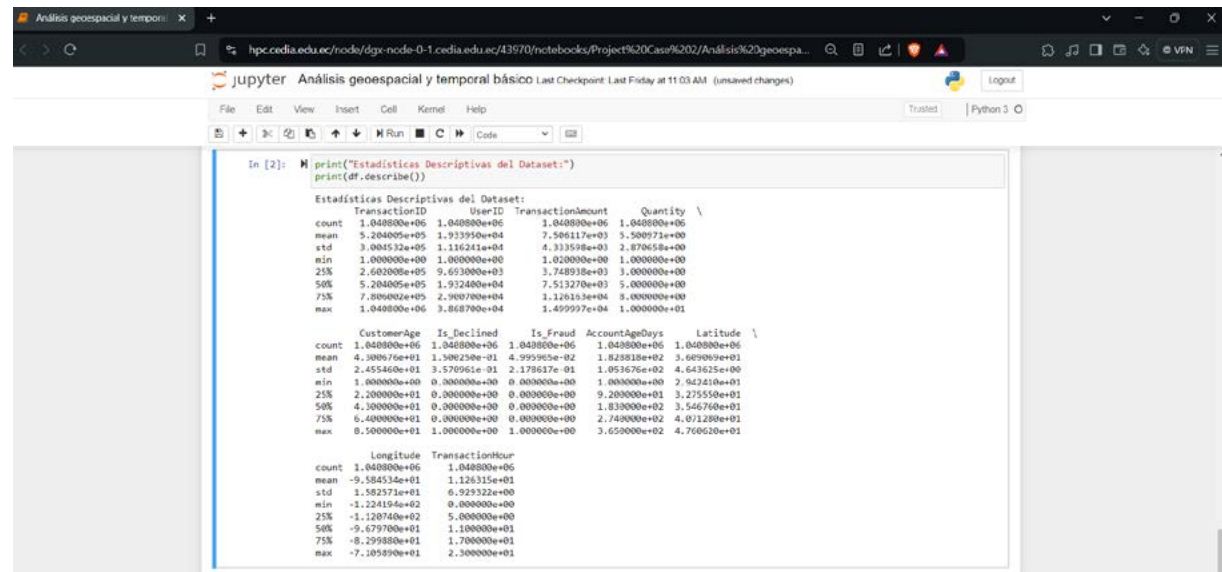


figura 32

## 2. Análisis Temporal: Distribución de transacciones por hora

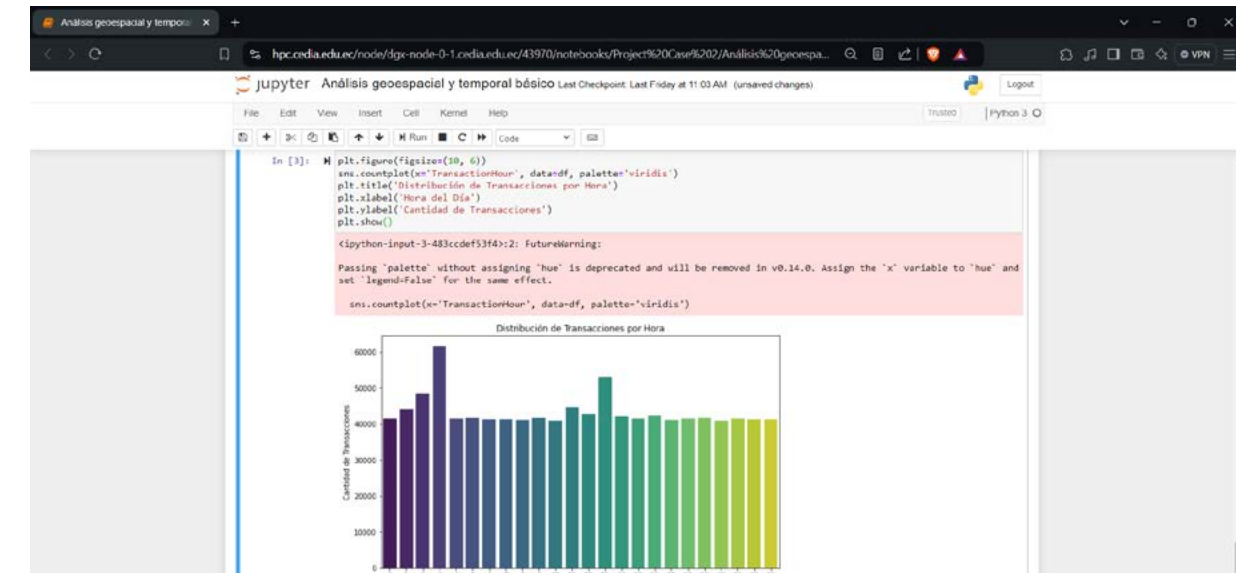


figura 33

## 3. Análisis Geoespacial: Mapa de calor de las transacciones

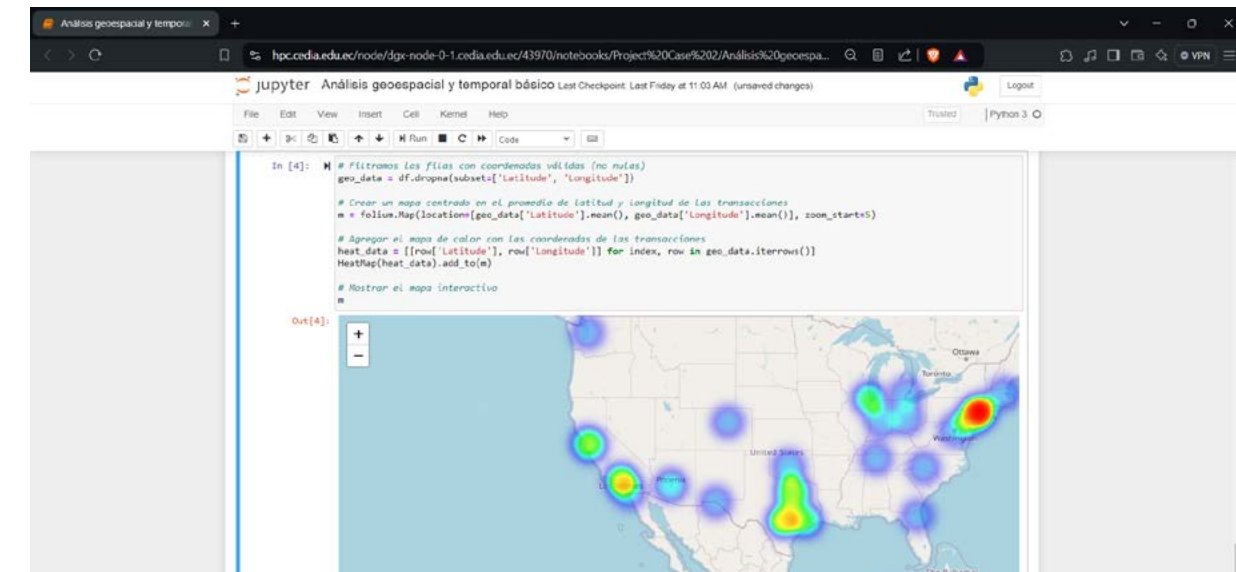


figura 34

#### 4. Análisis de Fraudes: Comparación de transacciones fraudulentas y no fraudulentas por ubicación

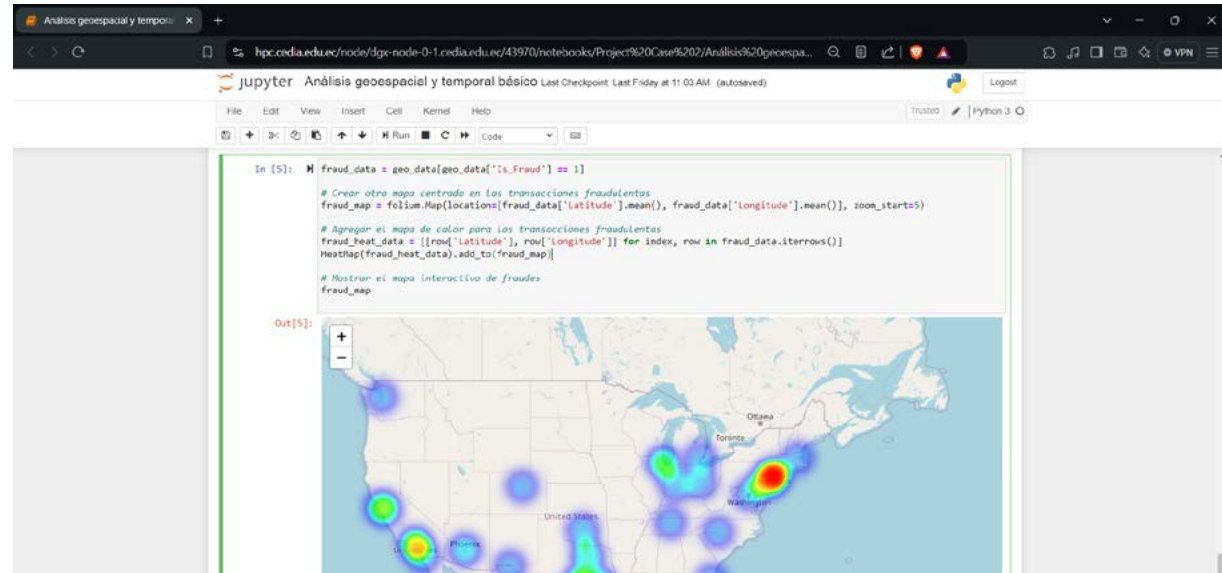


figura 35

#### 5. Análisis Temporal de Transacciones:

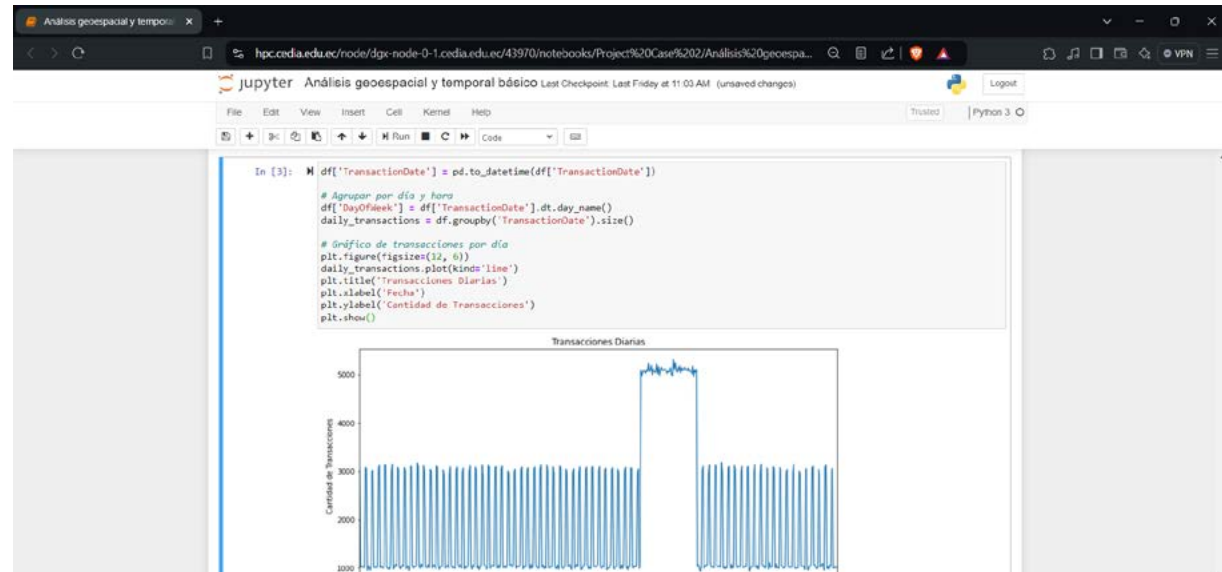


figura 36

#### 6. Análisis de Correlación

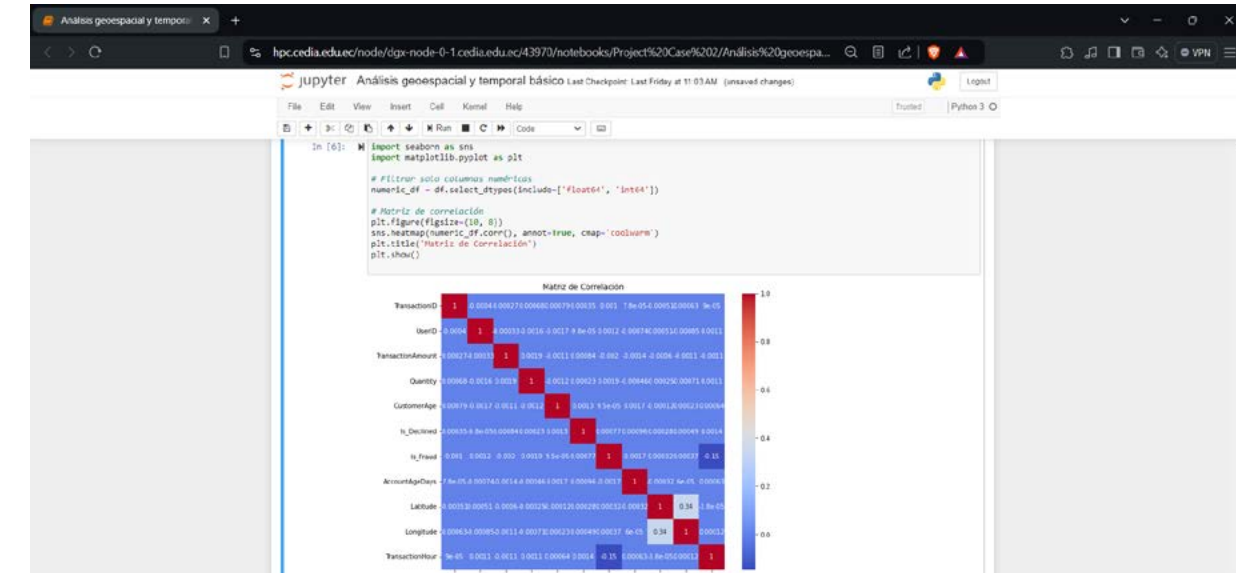


figura 37

#### 7. Análisis de Comportamiento del Usuario

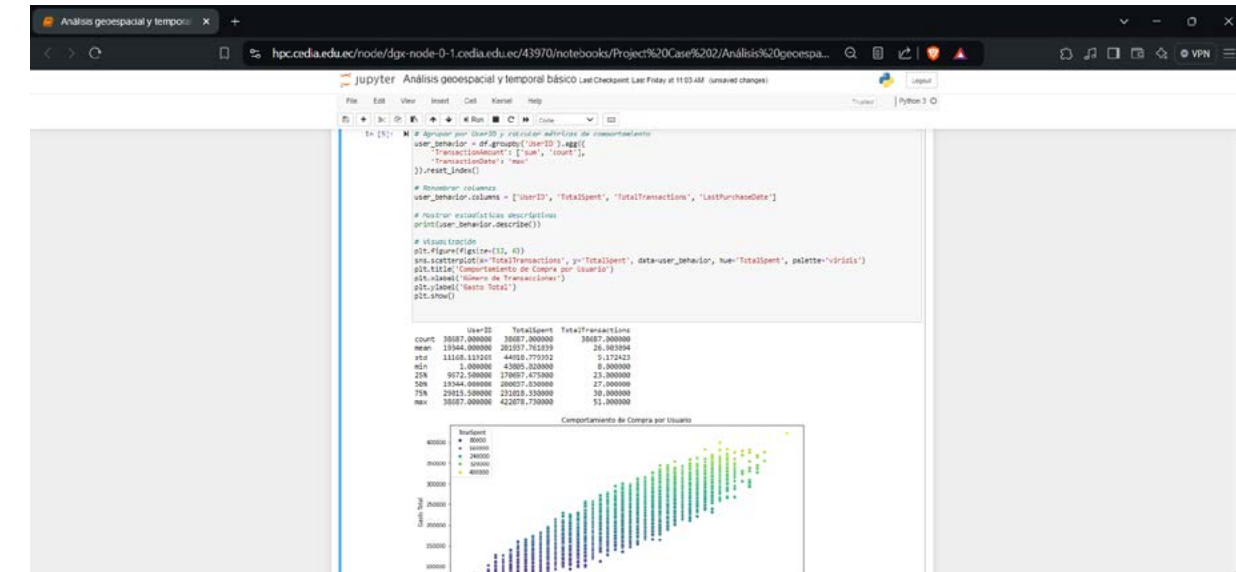


figura 38





## 8. Análisis del Método de Pago:

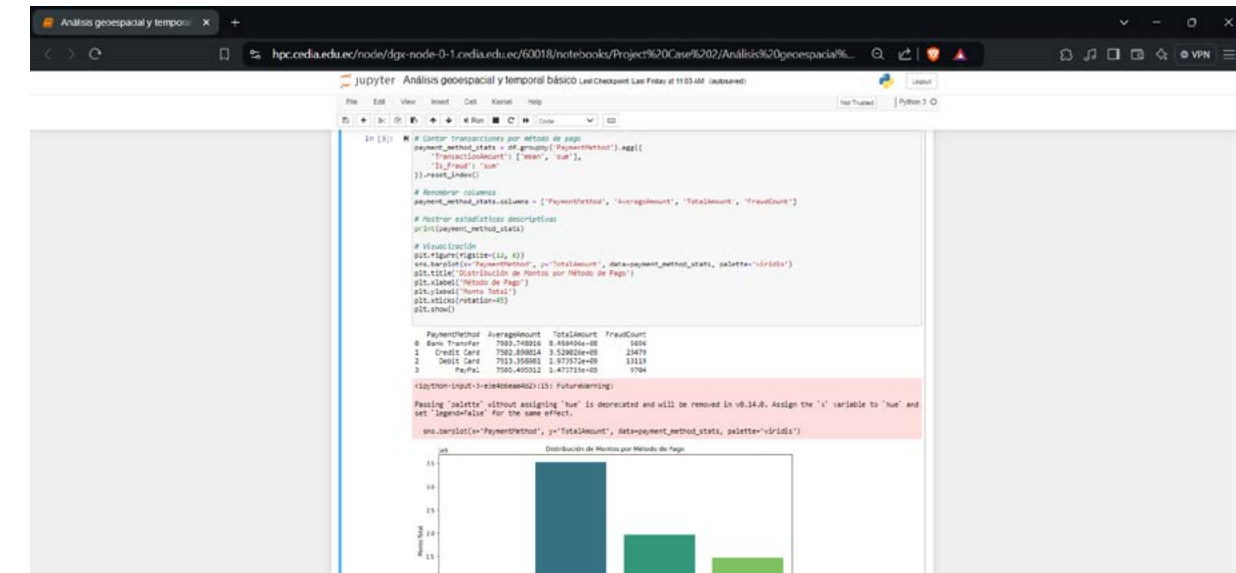


figura 39

## 9. Segmentación de Clientes Basada en el Monto de Transacción y Edad:

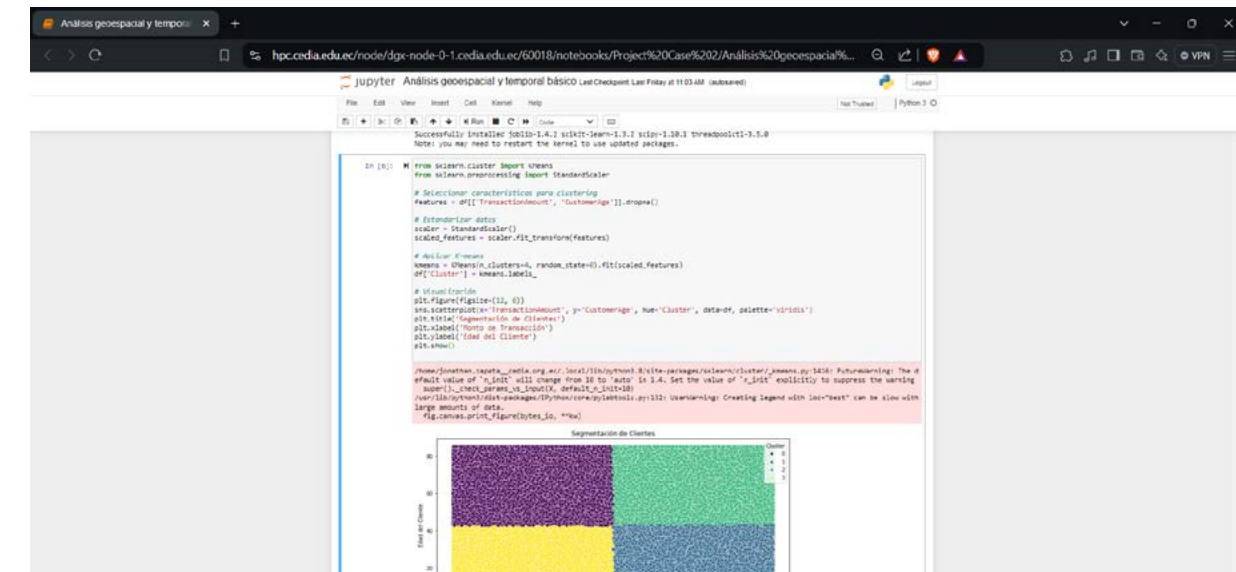


figura 40



## 10. Análisis de la Ubicación

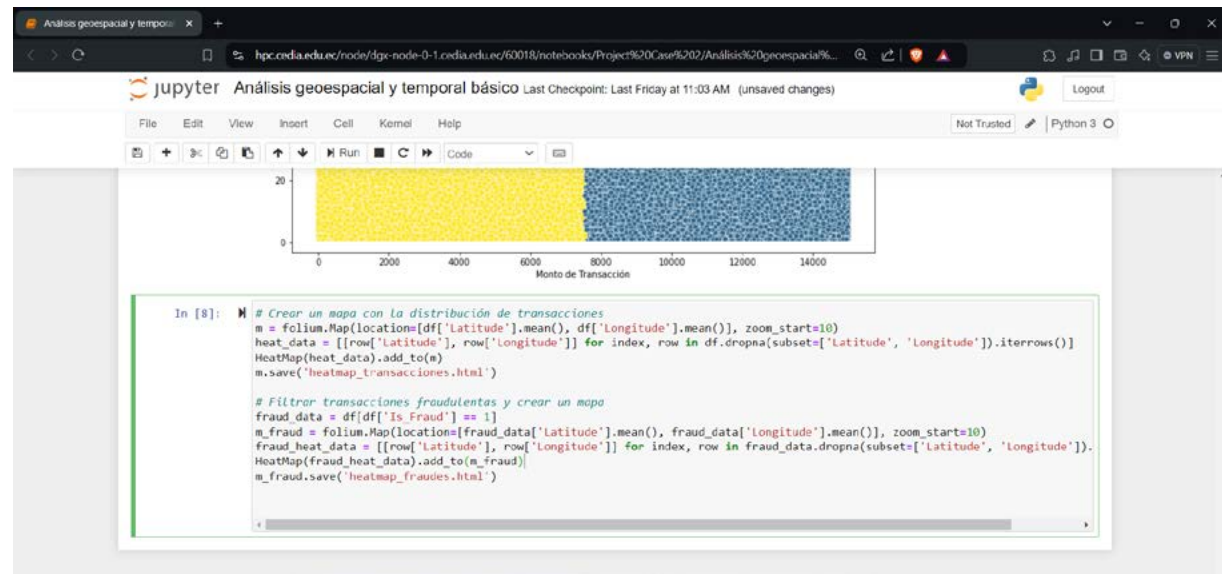


figura 41



figura 42

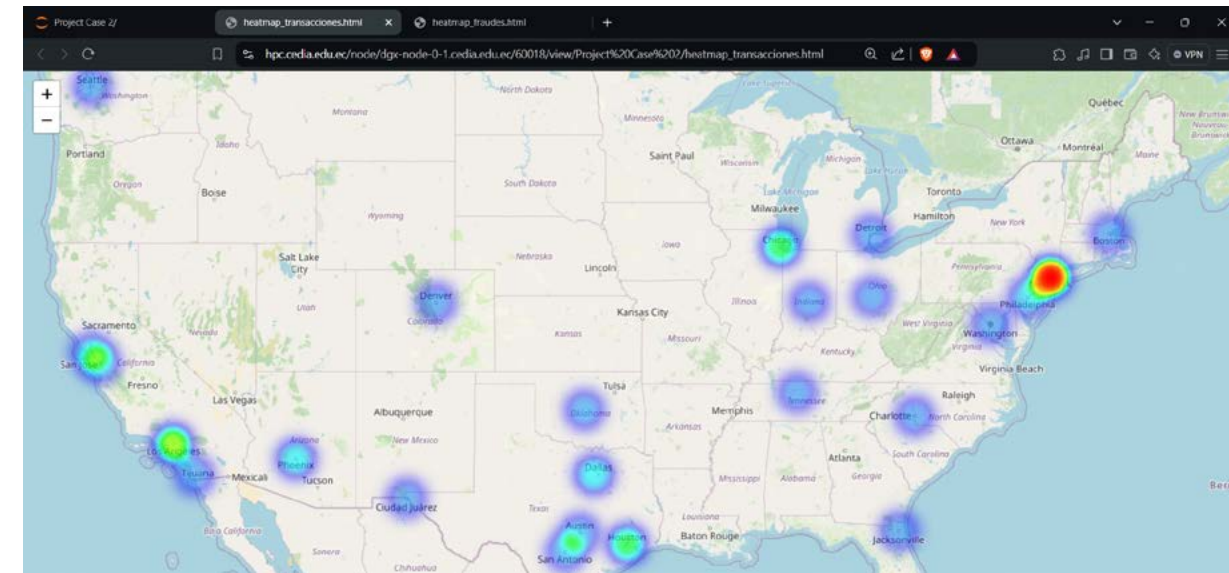


figura 43

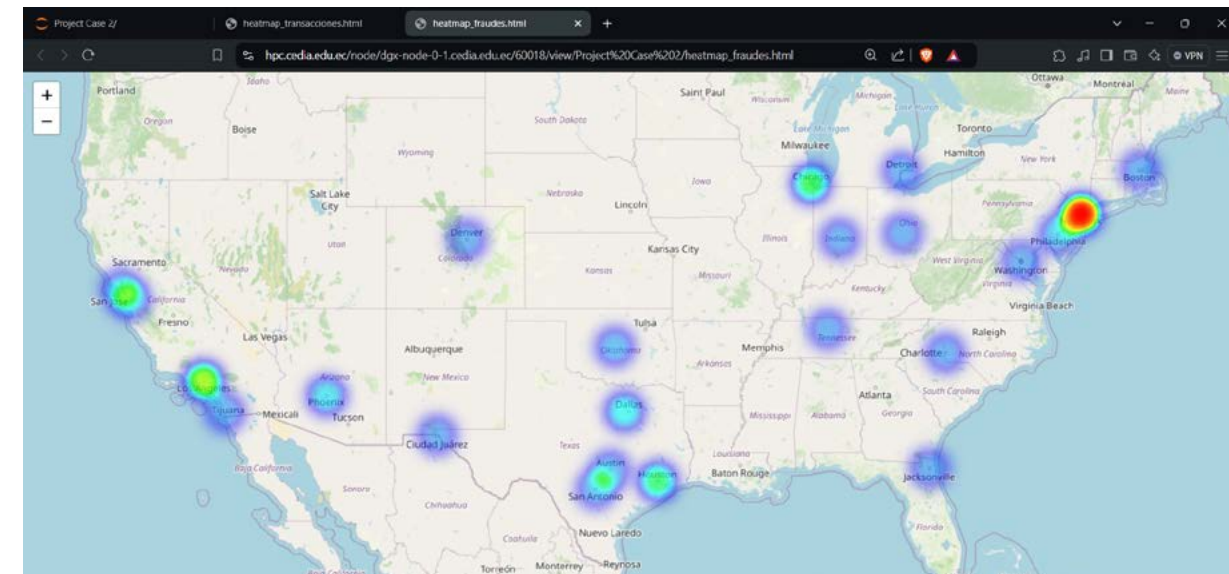


figura 44



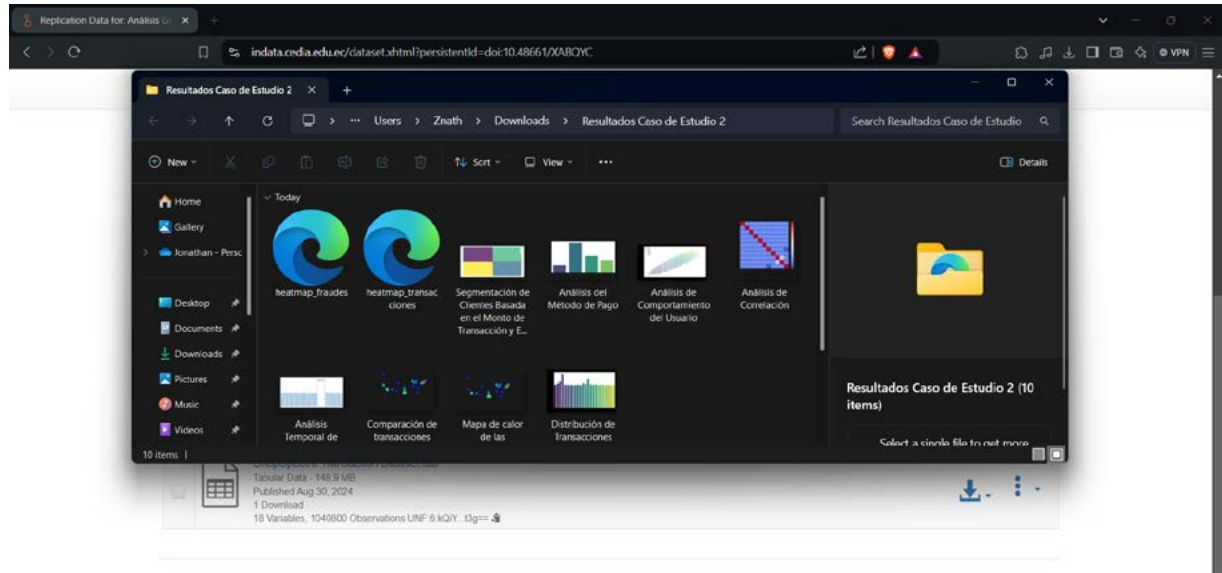


figura 45

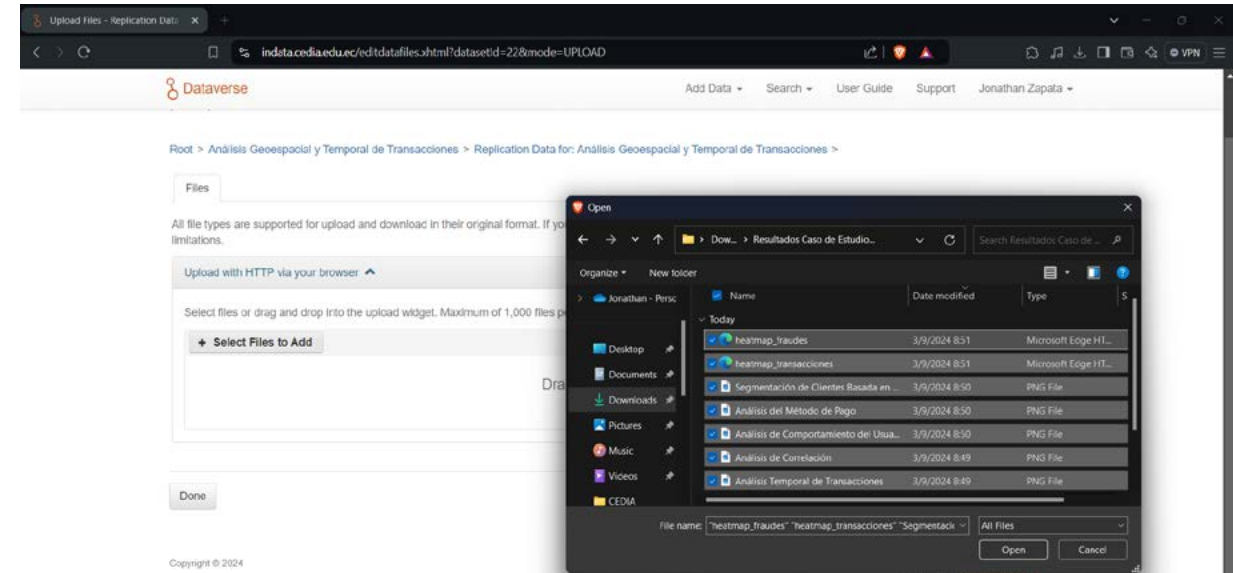


figura 47

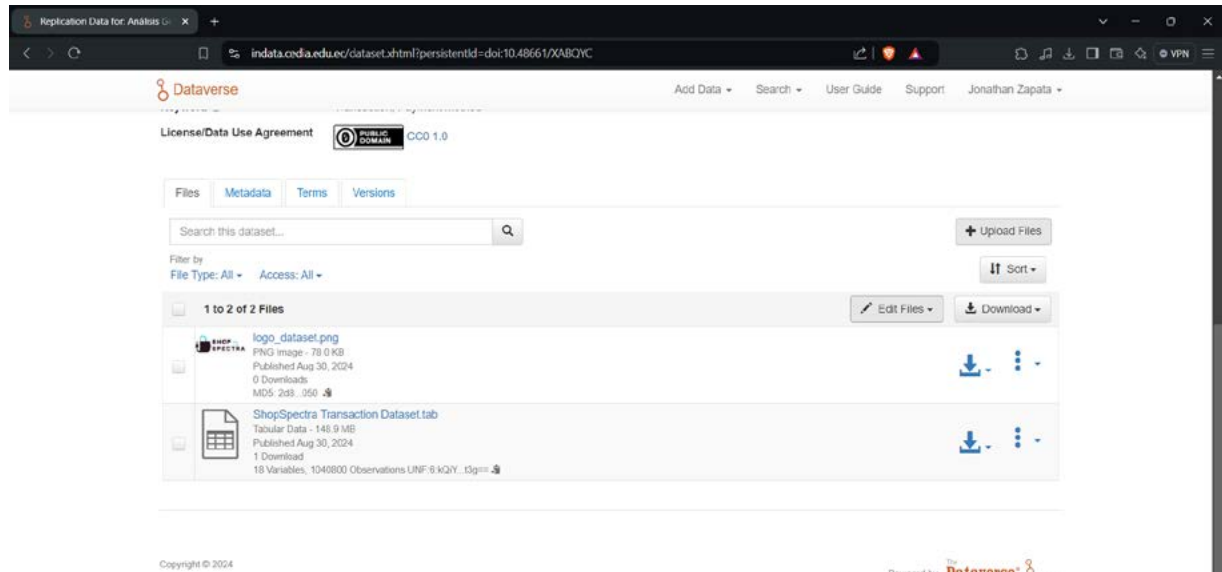


figura 46

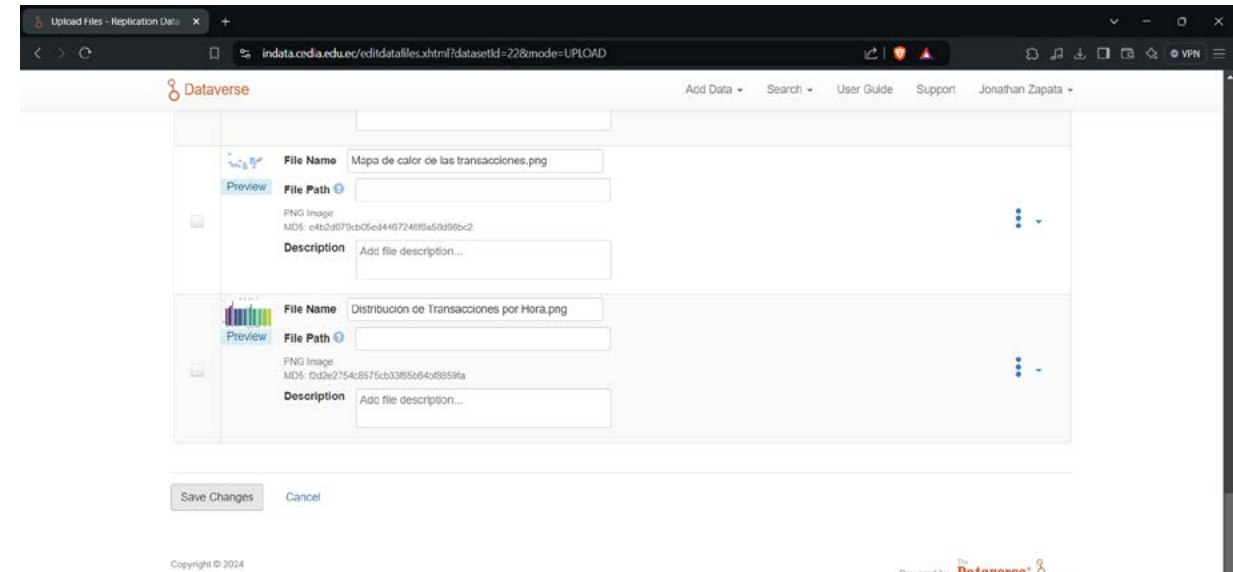


figura 48

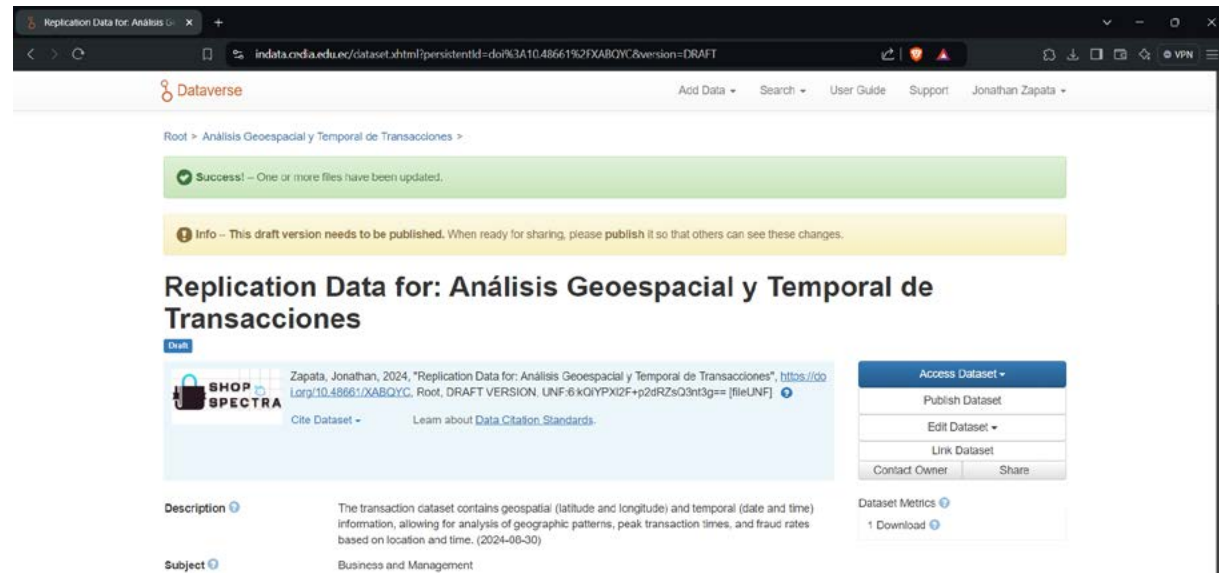


figura 49

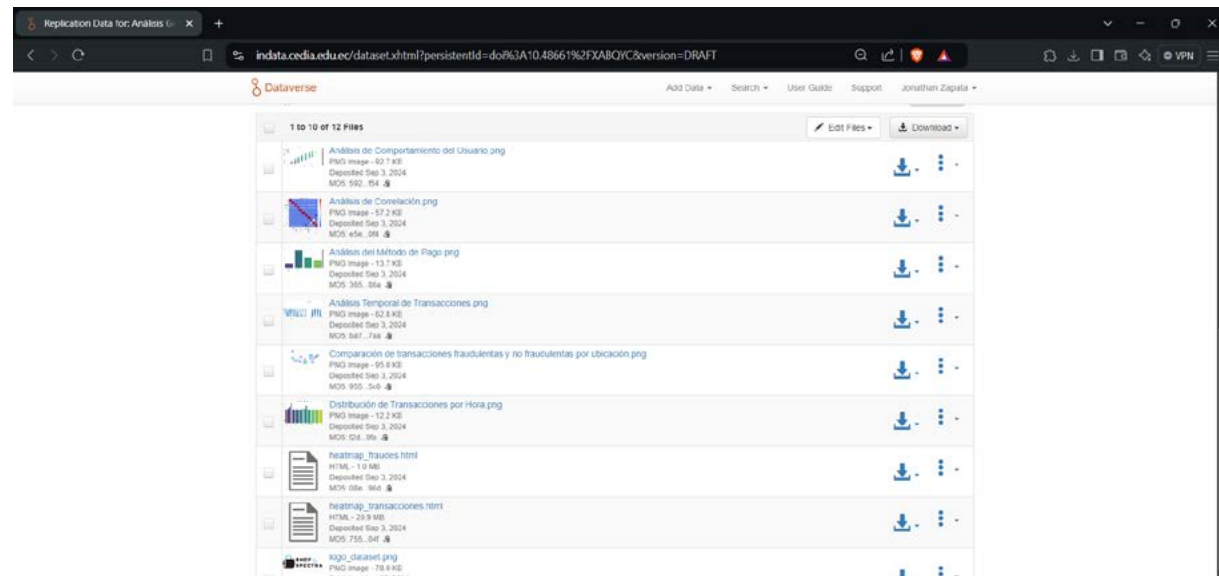


figura 50



#### 4. Análisis de Fraudes: Comparación de transacciones fraudulentas y no fraudulentas por ubicación

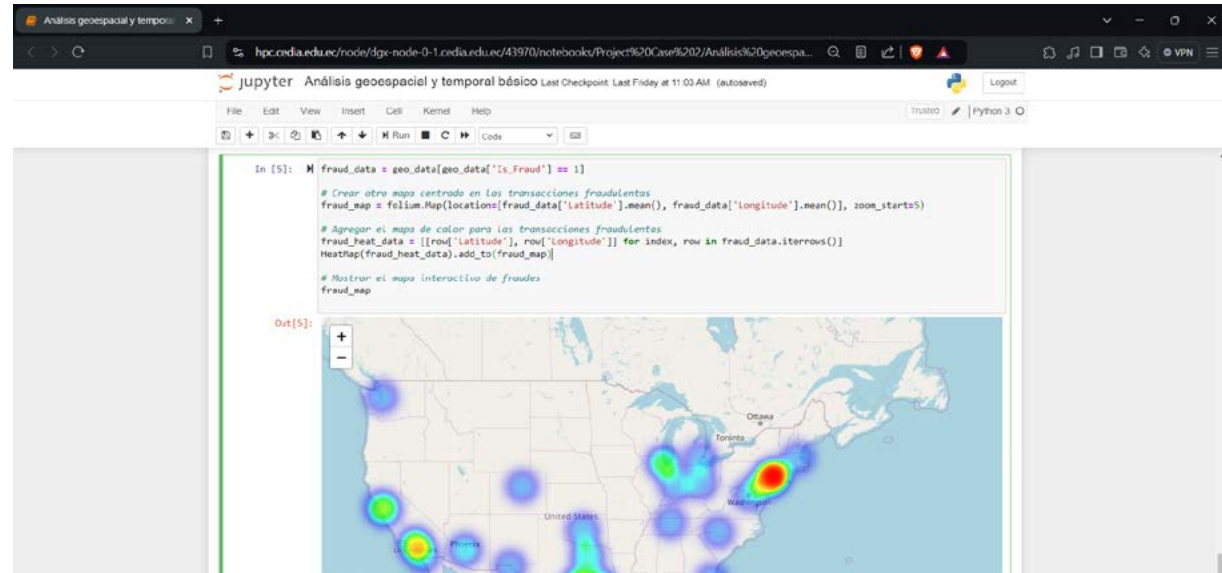


figura 35

#### 5. Análisis Temporal de Transacciones:

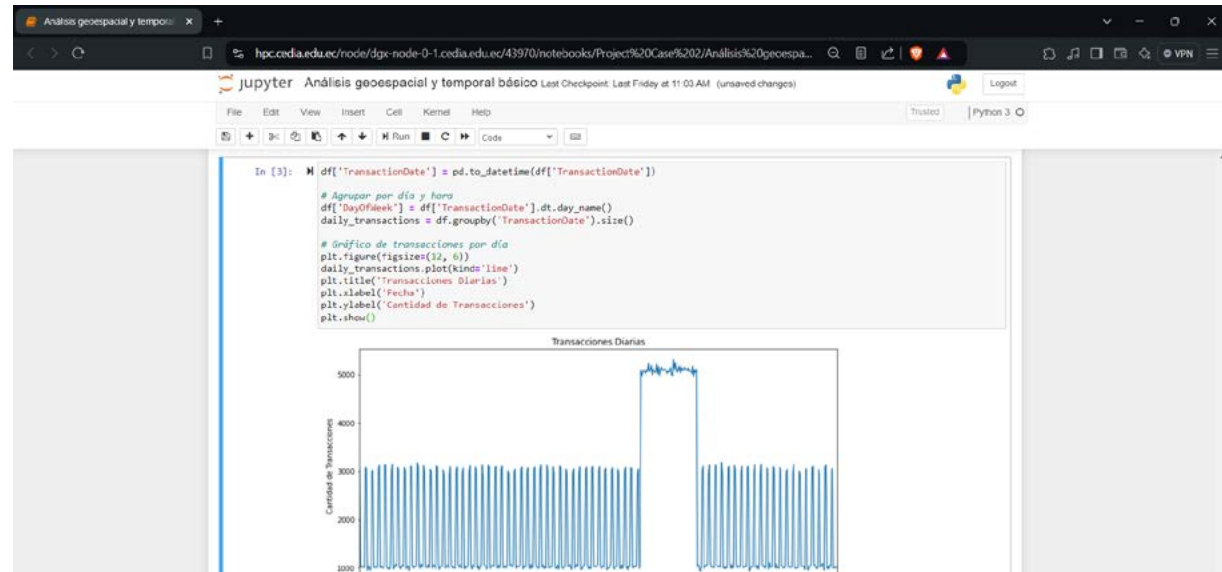


figura 36

#### 6. Análisis de Correlación

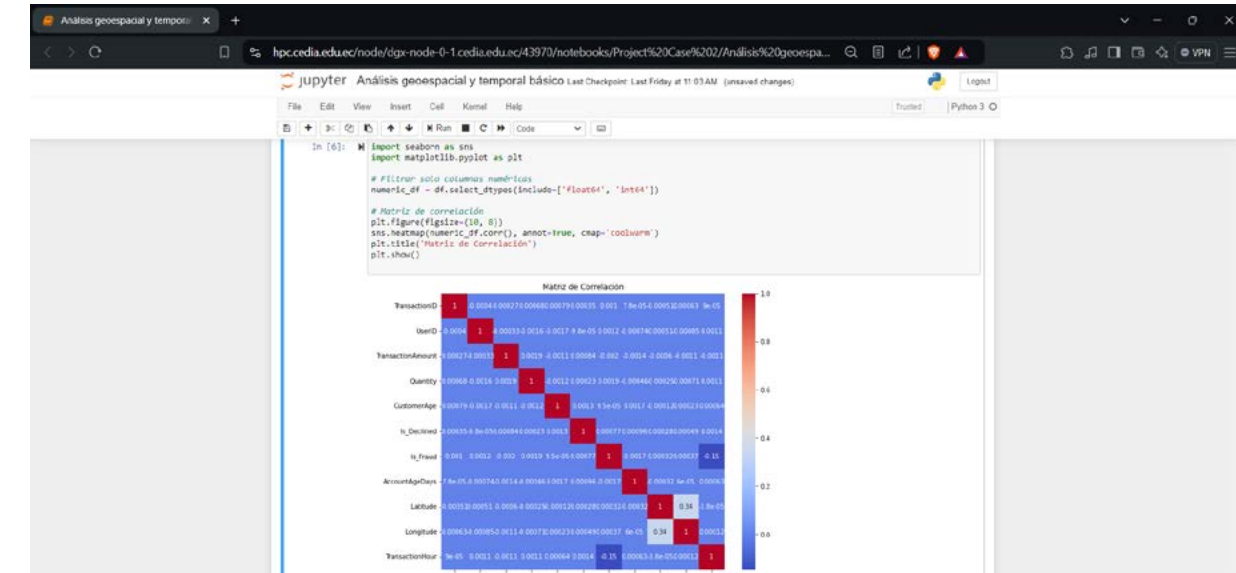


figura 37

#### 7. Análisis de Comportamiento del Usuario

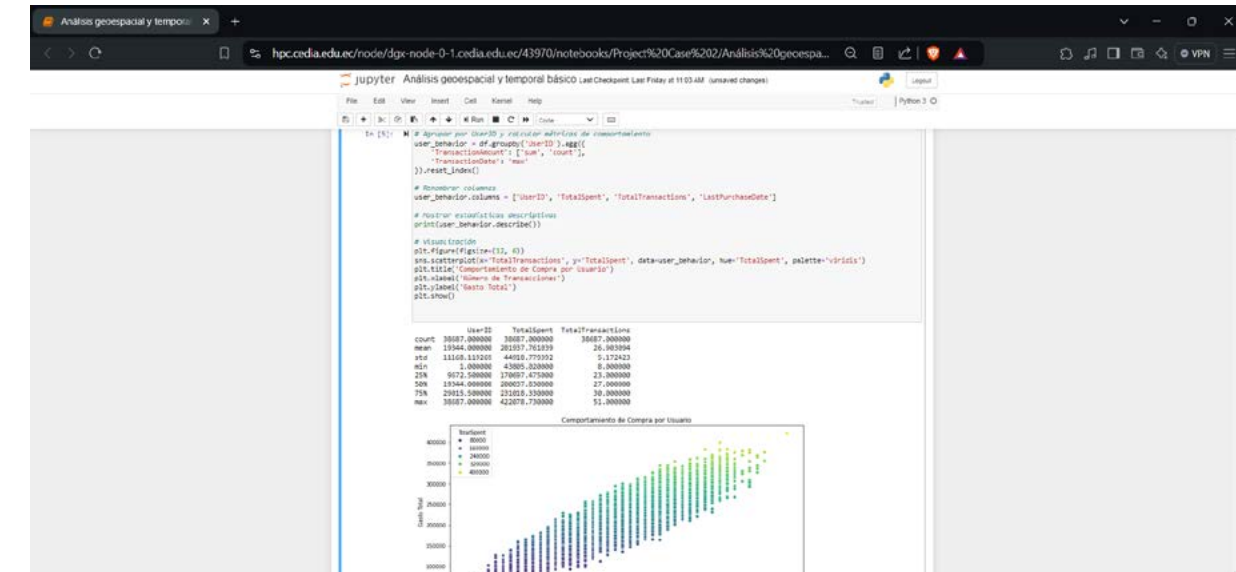
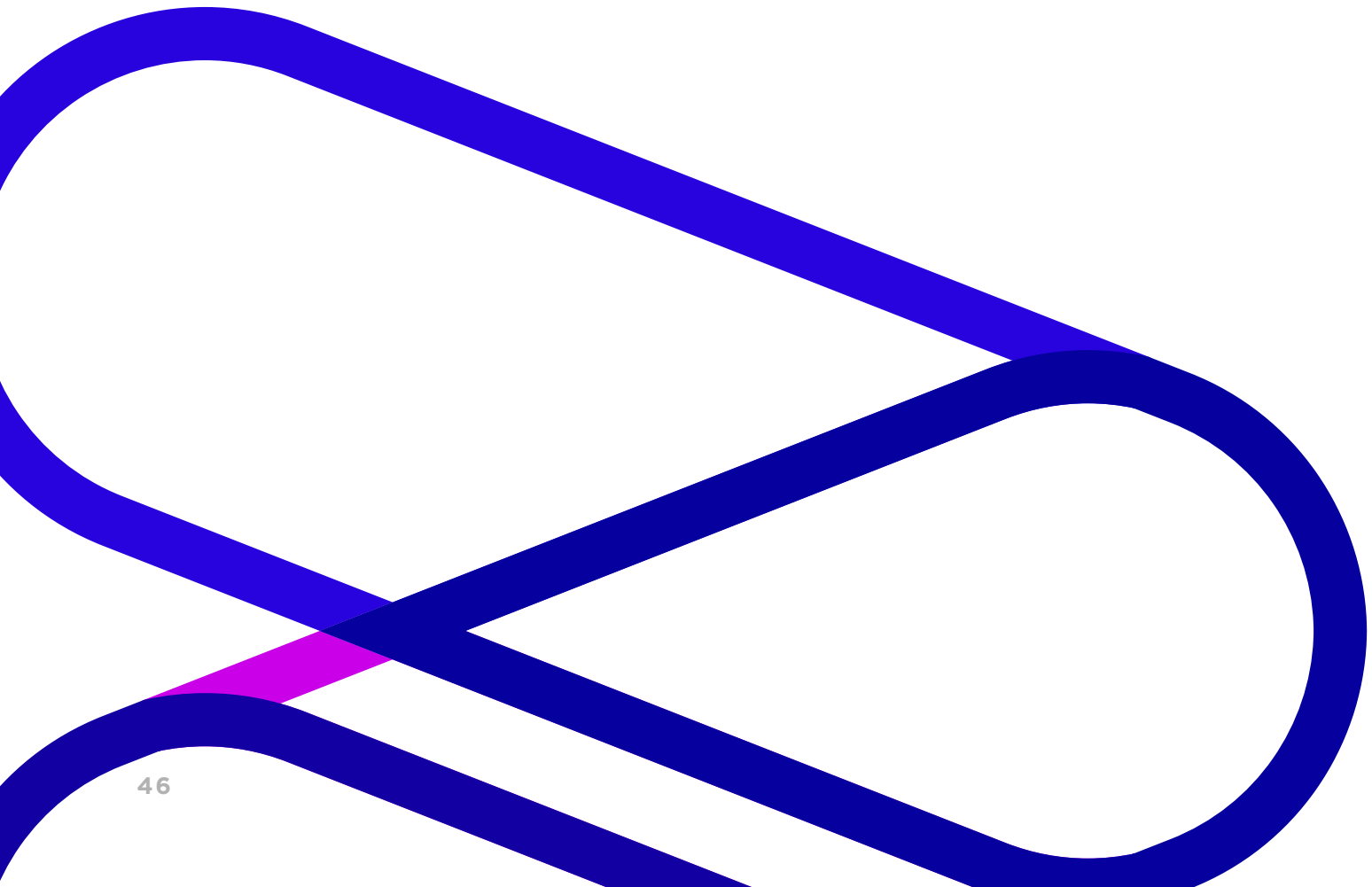


figura 38



1.4

## USO DE LA HERRAMIENTA GITLAB



Procederemos con la creacion del repositorio.

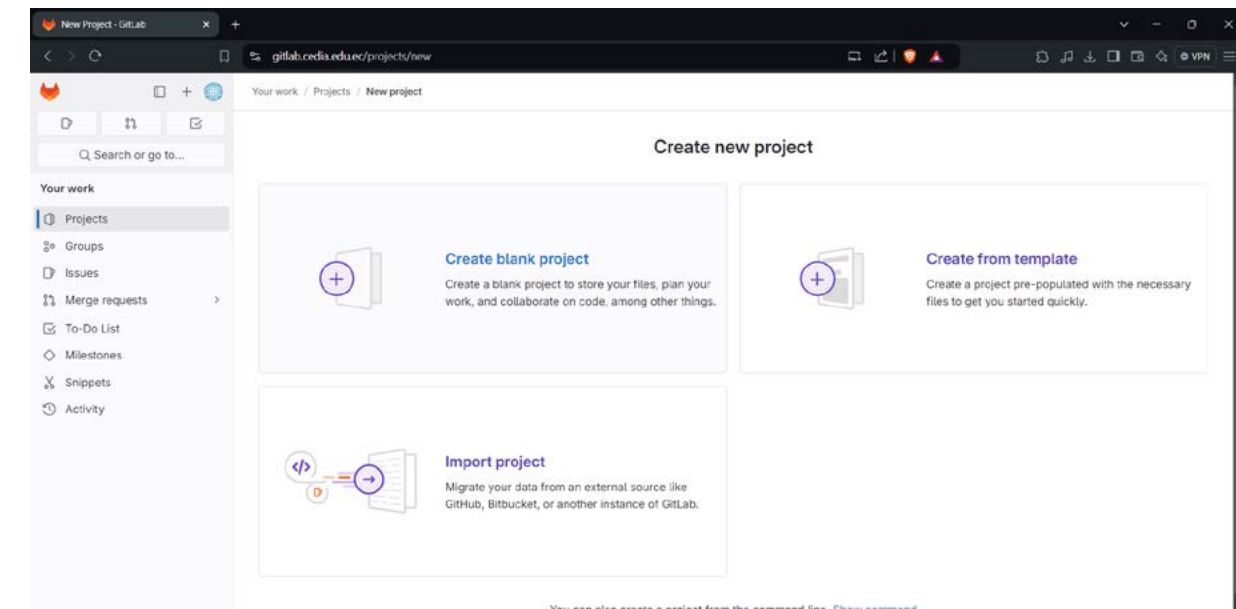


figura 51

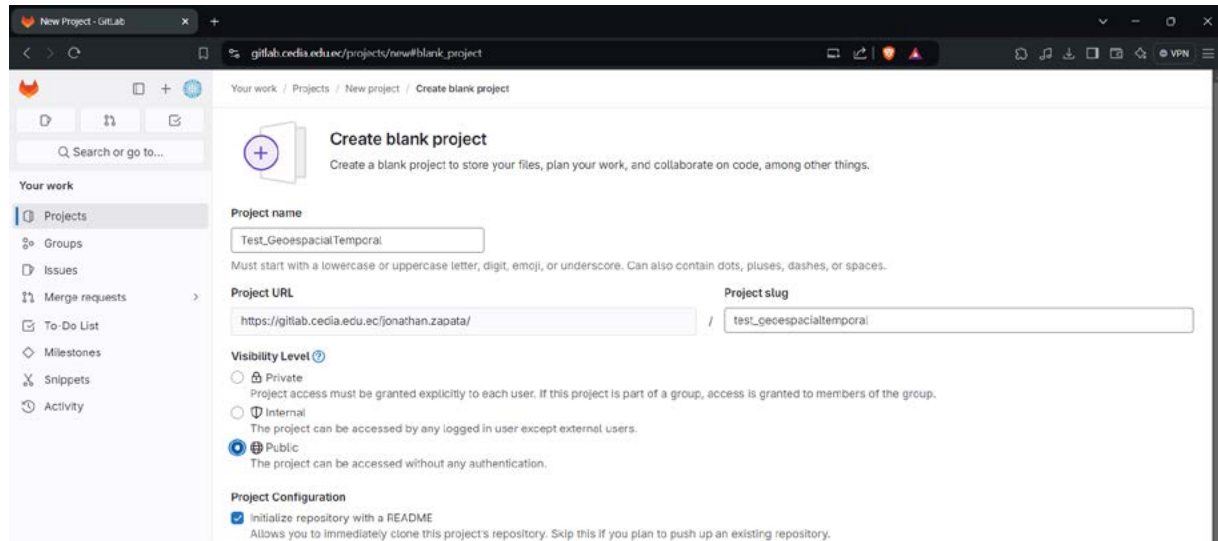


figura 52

Además, al ejecutar estos scripts de desarrollo en Python, es fundamental capturarlos o guardarlos adecuadamente para facilitar la colaboración en equipo y permitir el mantenimiento o las mejoras continuas del código. Por ello, se utiliza la plataforma GitLab para almacenar todo en un único repositorio, asegurando así una gestión eficiente y un acceso centralizado a los scripts.

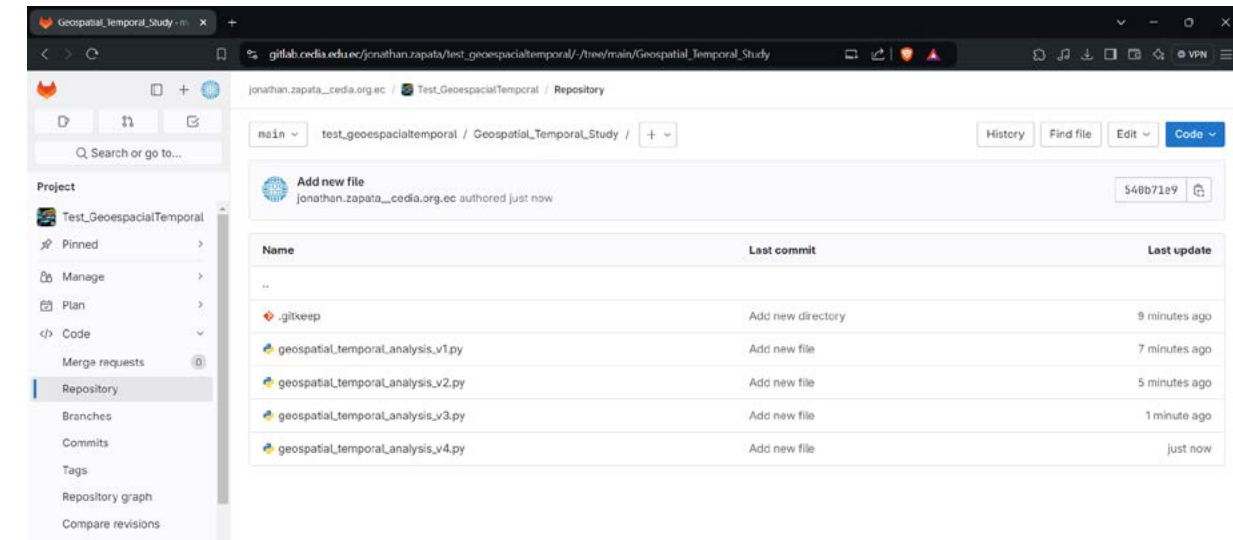


figura 54

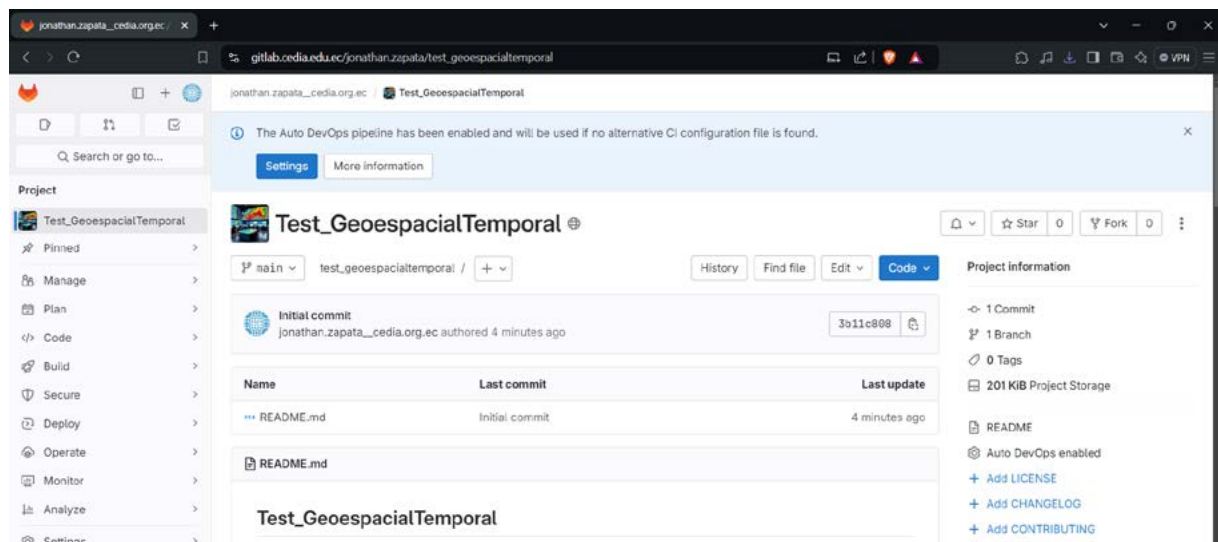


figura 53

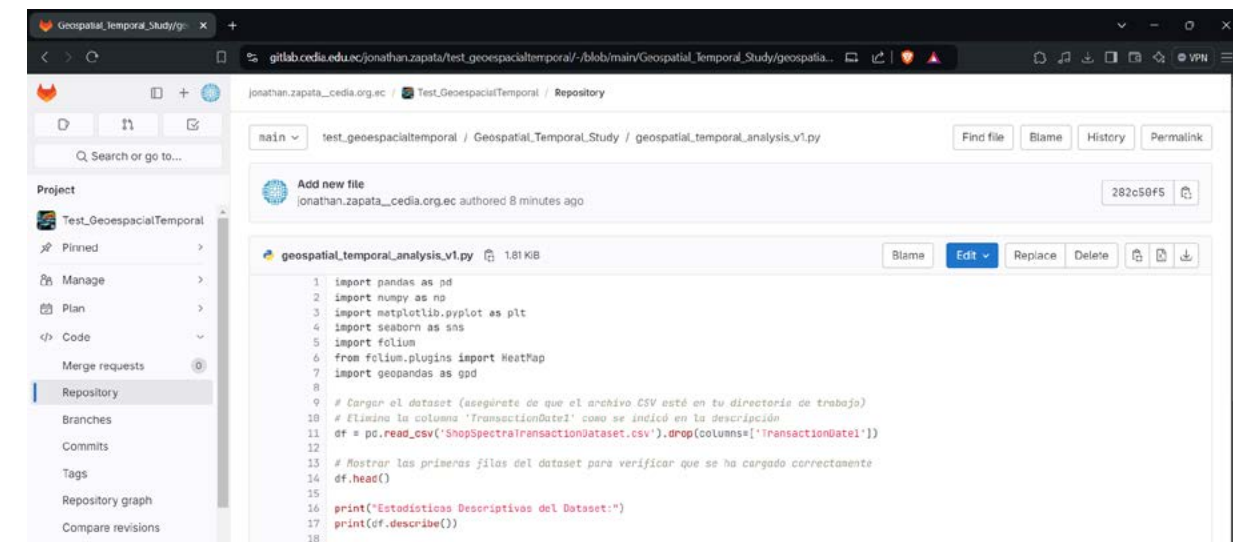


figura 55

Para implementar la codificación en la plataforma Git, es fundamental considerar la colaboración y el trabajo en equipo, permitiendo que todos los integrantes del proyecto puedan contribuir y aportar de manera efectiva a la investigación.

Vamos a la **invitación de colaboradores**.

Nos dirigimos a Manage >Members y finalmente Invite Members . A continuación, invita a todos los colaboradores que participarán en el proyecto.

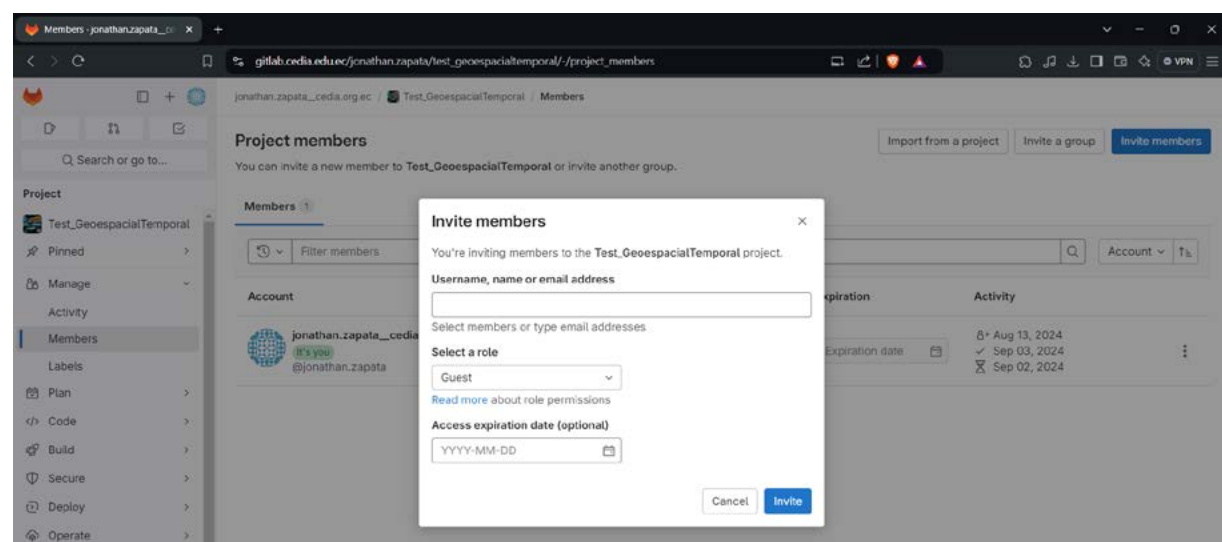


figura 56

Y aqui es muy importante el tipo de rol a designar que esta entre Guest, reporter, developer, maintainer y owner.

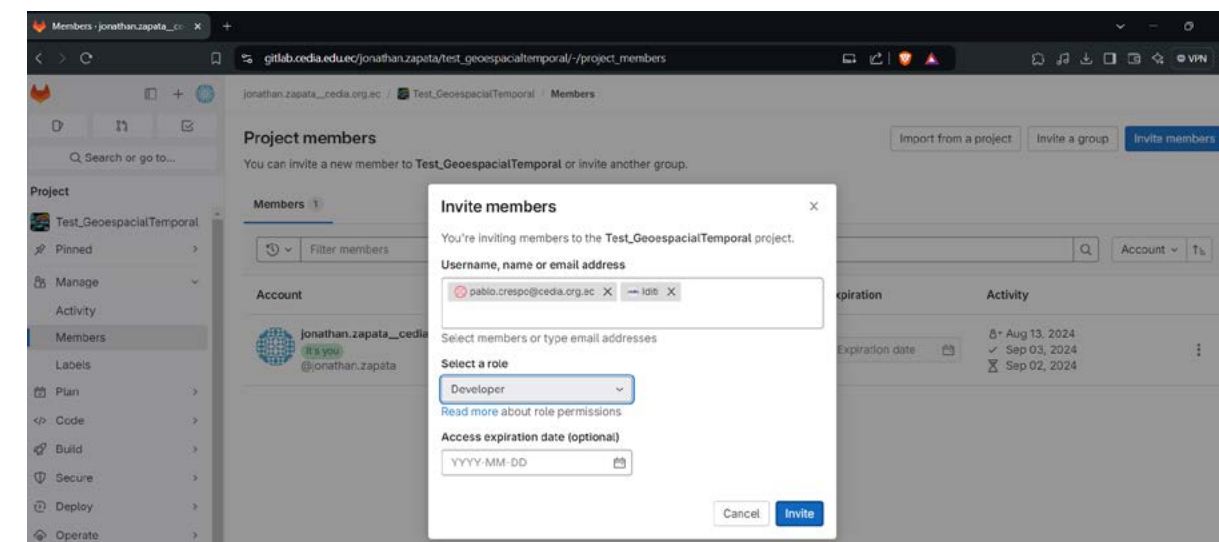


figura 57

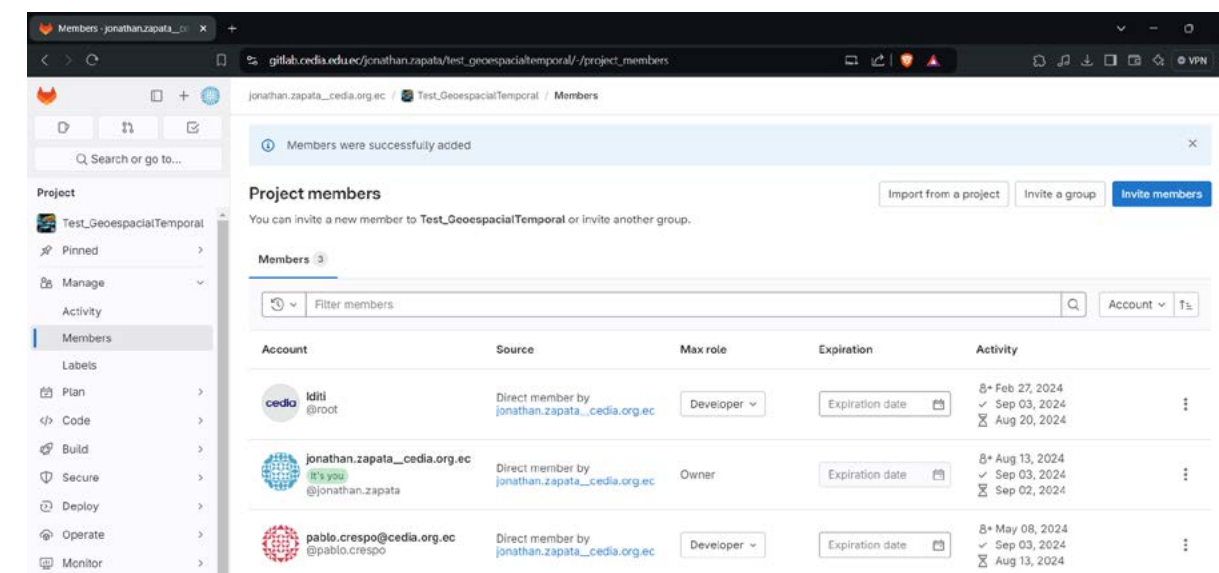


figura 58



Ahora procederemos a la asignación de permisos para cada colaborador en la rama principal. En GitHub, la rama principal (branch) representa la línea principal de desarrollo del código.

Nos dirigimos a Settings > Repository y configuramos los permisos para que tanto los mantenedores (maintainers), los desarrolladores (developers) e instance admin tengan acceso a la rama principal.

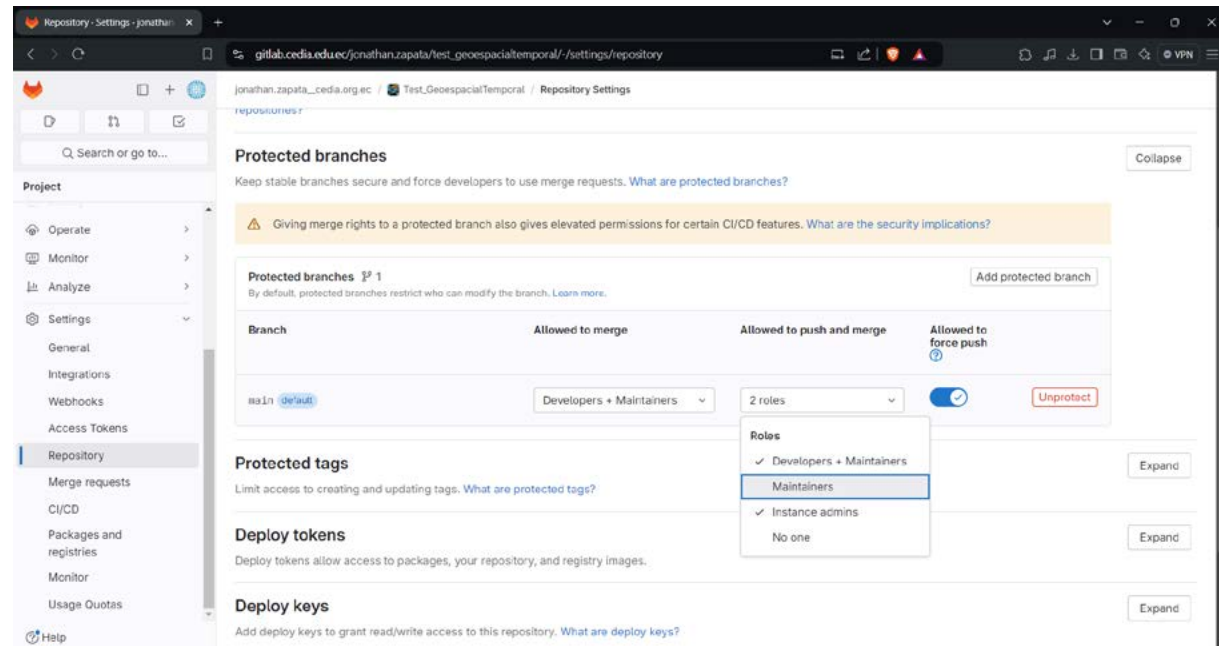


figura 59

Finalmente, para la revisión o la designación de un revisor en la investigación, se le puede asignar un token de acceso en caso de que el repositorio sea privado. Este token le permitirá ingresar al repositorio de Git, y se puede generar en la sección de Settings > Access Tokens.

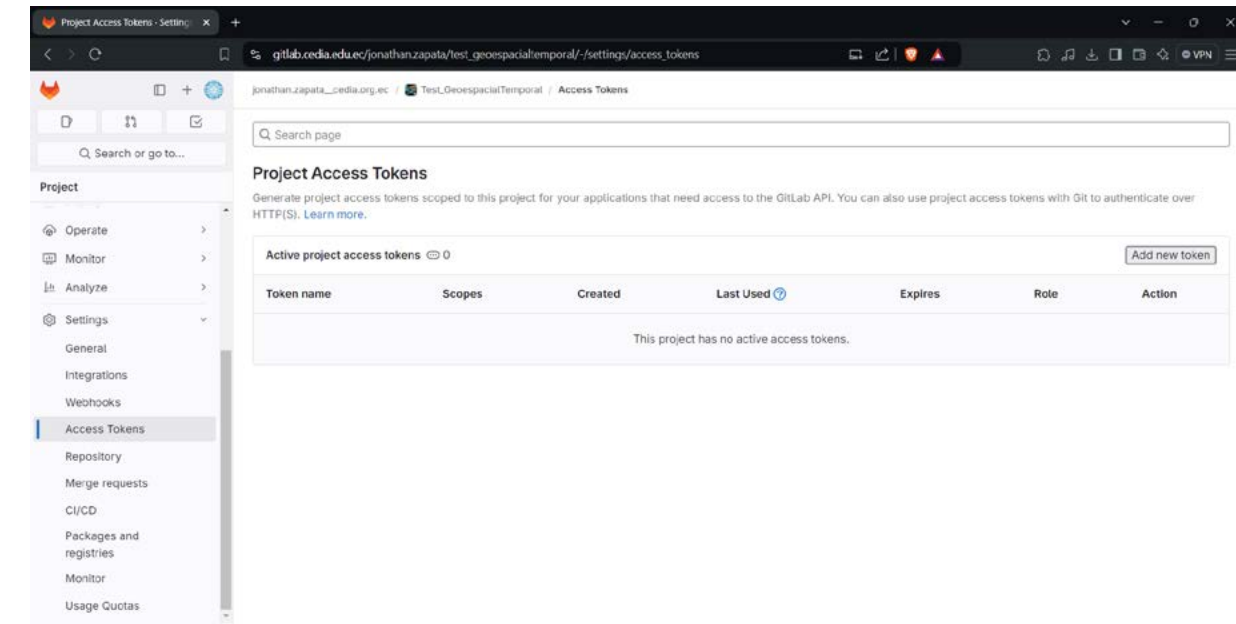


figura 60





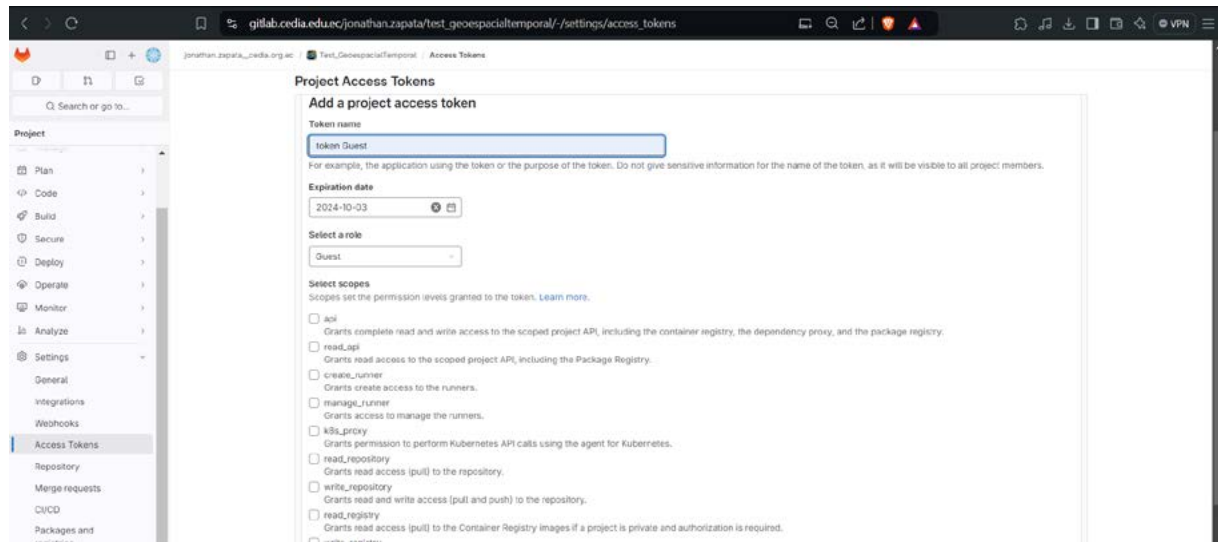


figura 61

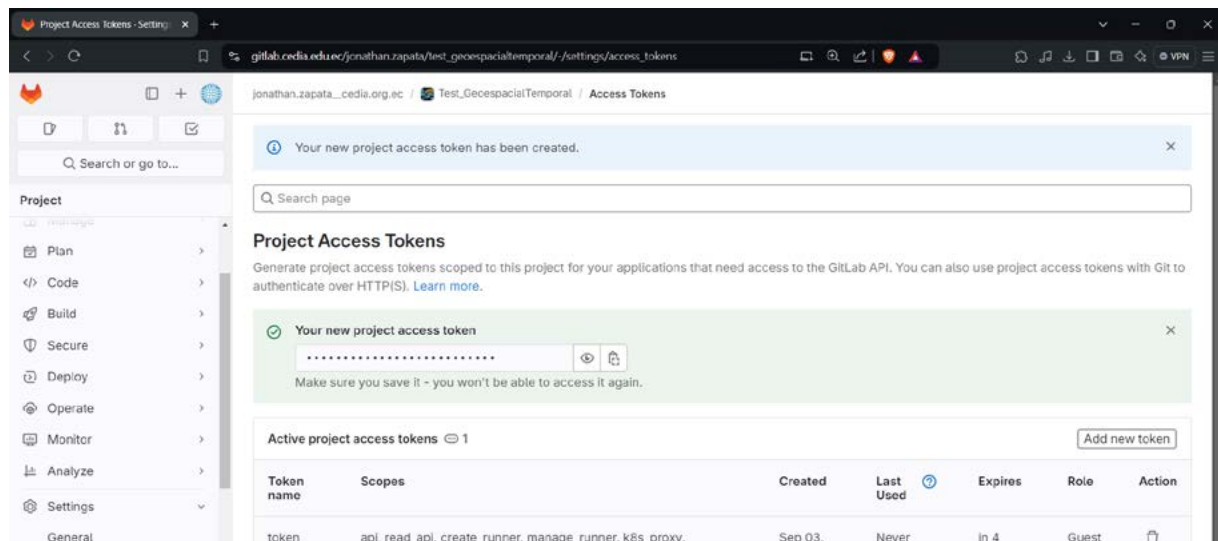


figura 62

Una vez se le entrega el token correspondiente, el revisor podrá clonar el repositorio y tener acceso al código.



# 1.5

## USO DE LA HERRAMIENTA MATTERMOST



Finalmente, se utiliza Mattermost como la herramienta principal de comunicación entre los participantes de la investigación. Esta plataforma permite coordinar reuniones para discutir los resultados y el avance del proyecto. Se configura un canal con el nombre del proyecto o una palabra clave relevante, como “Geospatial\_Temporal\_Study\_2024”, y se invita a los colaboradores involucrados en la investigación para facilitar la colaboración y el intercambio de ideas en tiempo real.

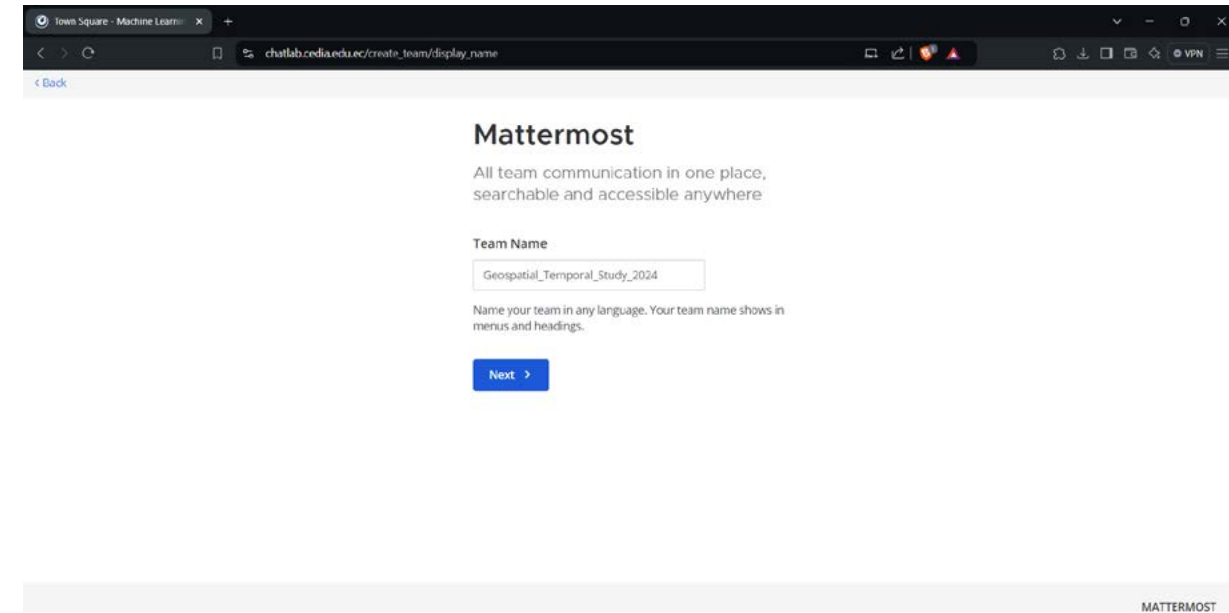


figura 63

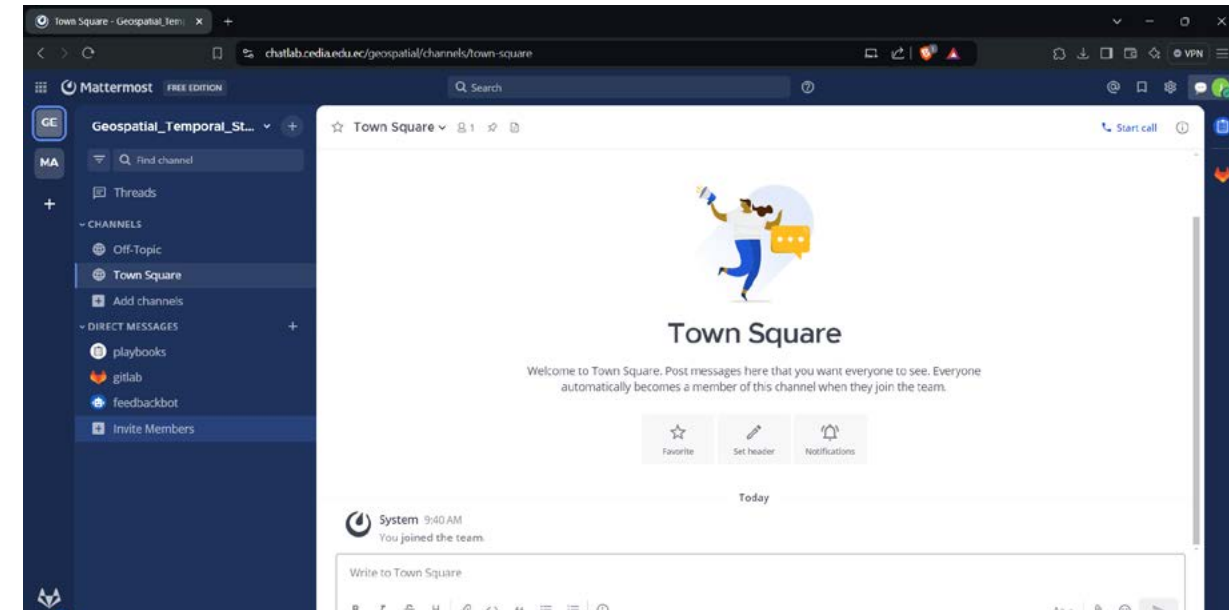


figura 64



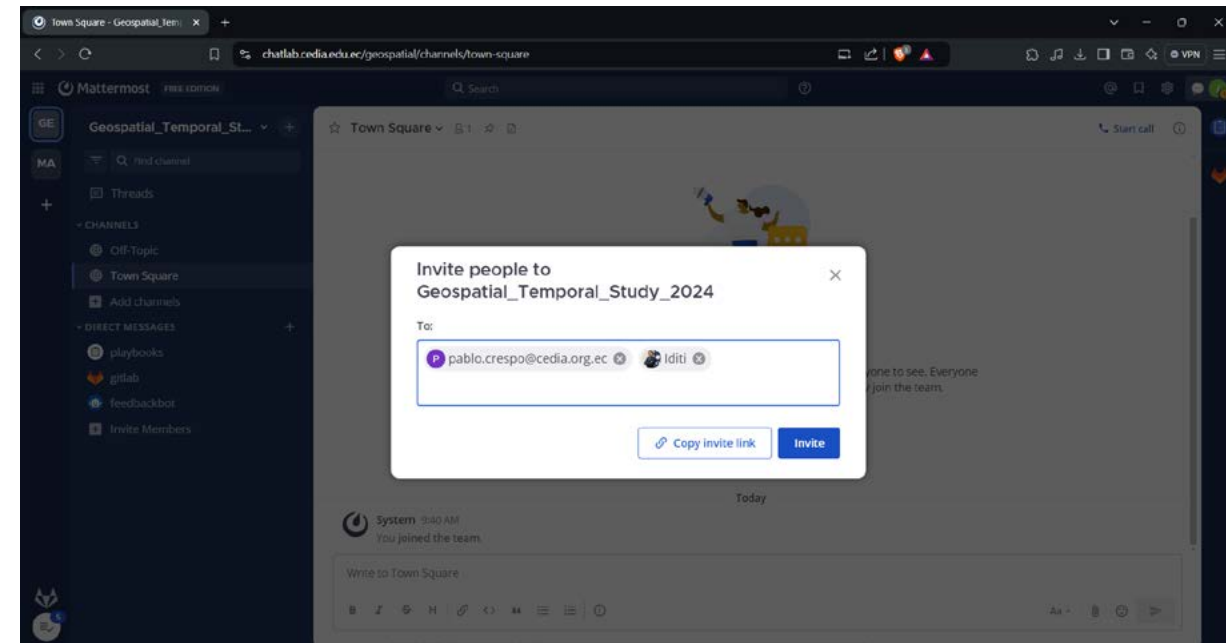


figura 65

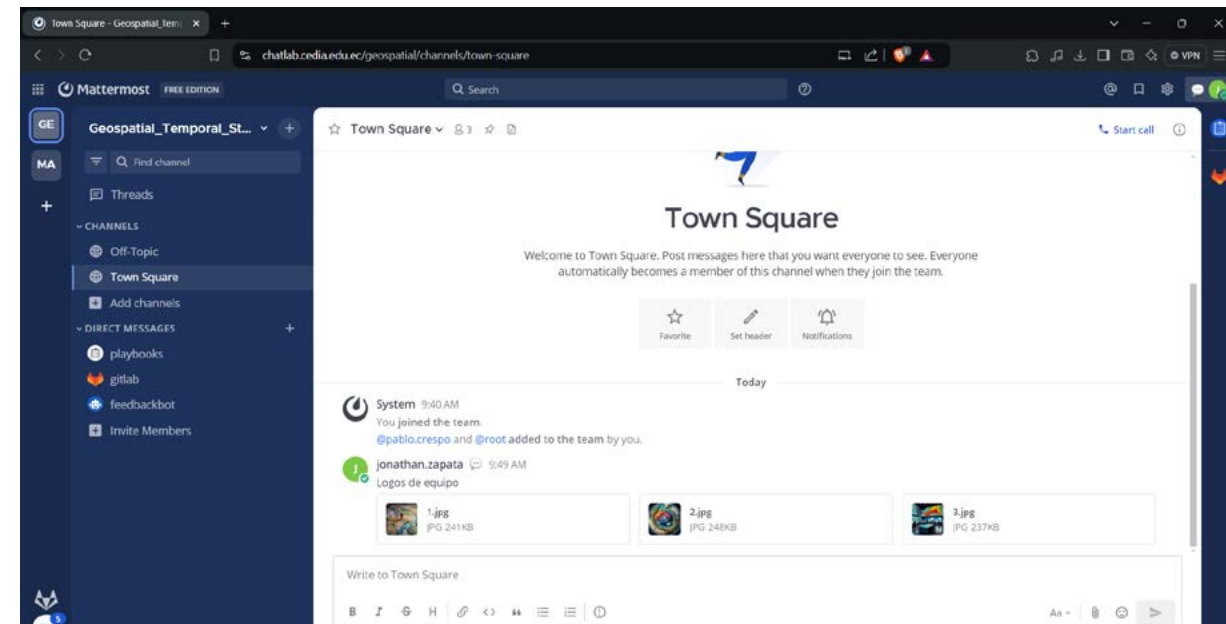


figura 66

Mattermost ofrece una amplia gama de herramientas, como la personalización del perfil, notificaciones, configuración de estado, y la capacidad de identificar y unirse a canales públicos o privados, así como enviar mensajes directos.

Además, es posible compartir fotos, links, código en formato Python, lo que permite a los investigadores visualizar el código que se está utilizando y evaluar posibles cambios o mejoras de manera más efectiva.

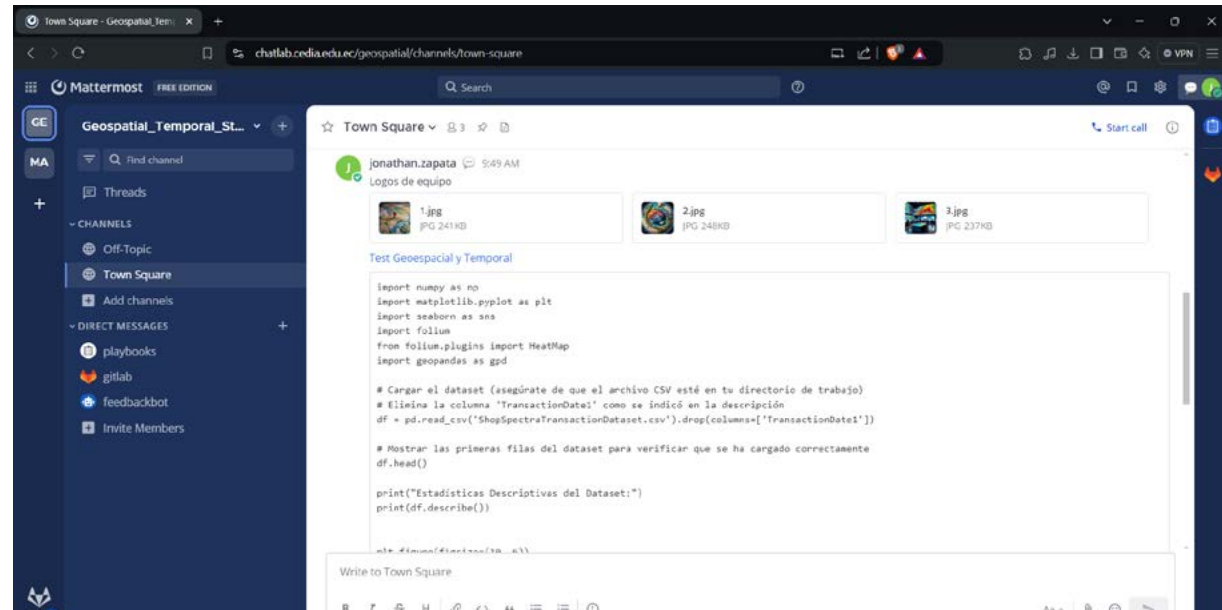


figura 67

## Creacion de PLAYBOOKS

Colección de procedimientos y directrices documentadas que guían a los equipos en la gestión de tareas o procesos repetitivos

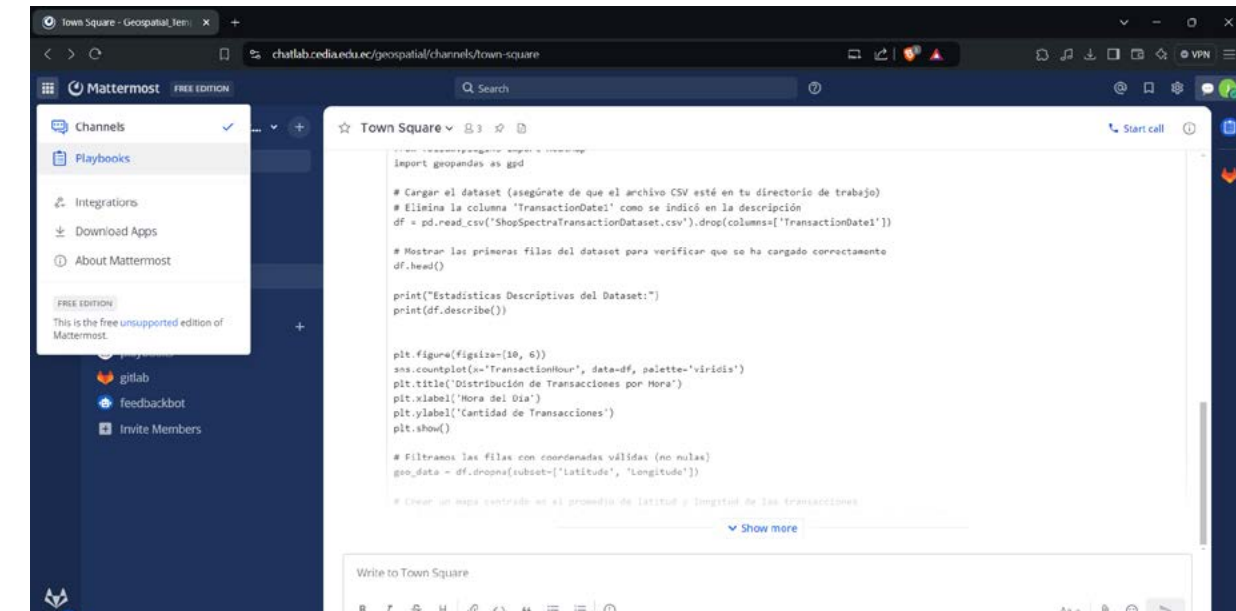


figura 68



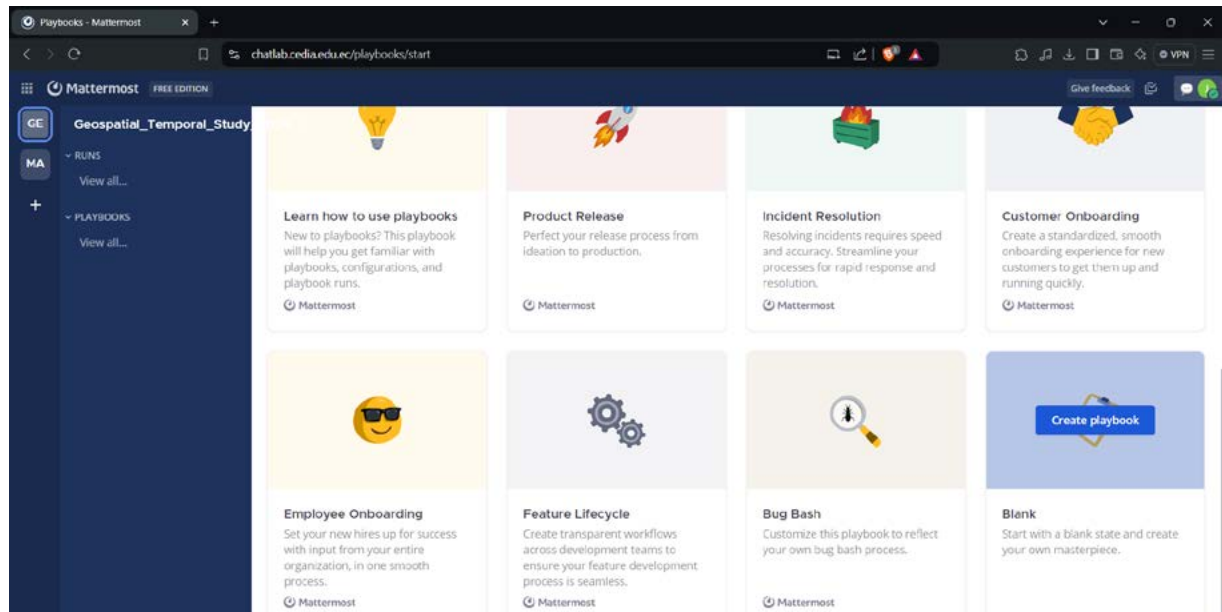


figura 69

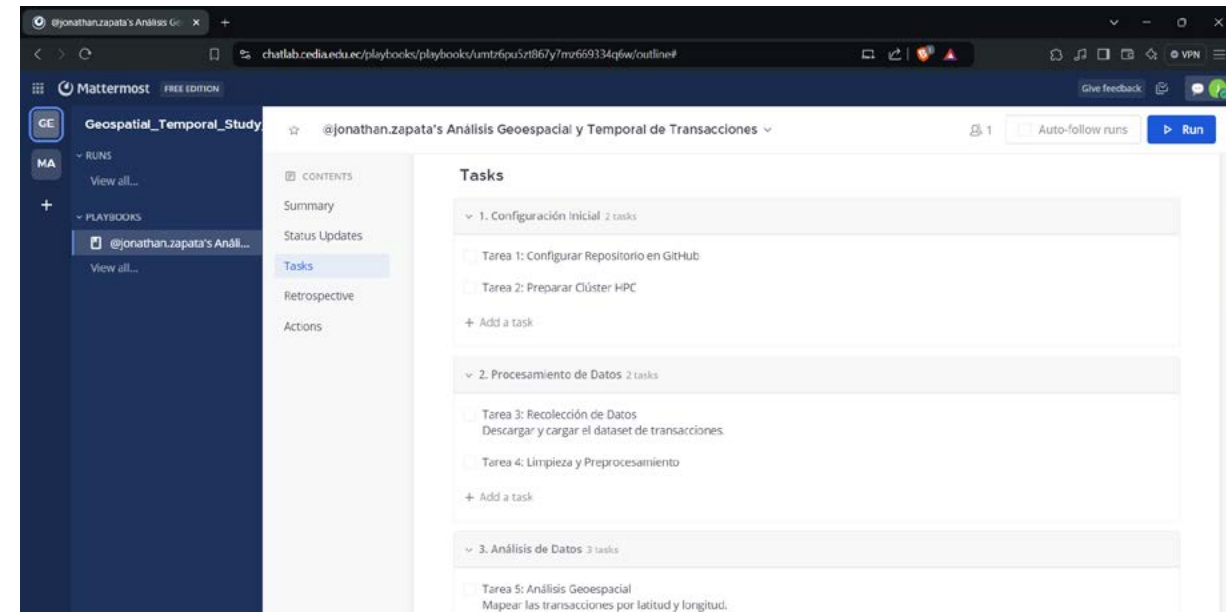


figura 71

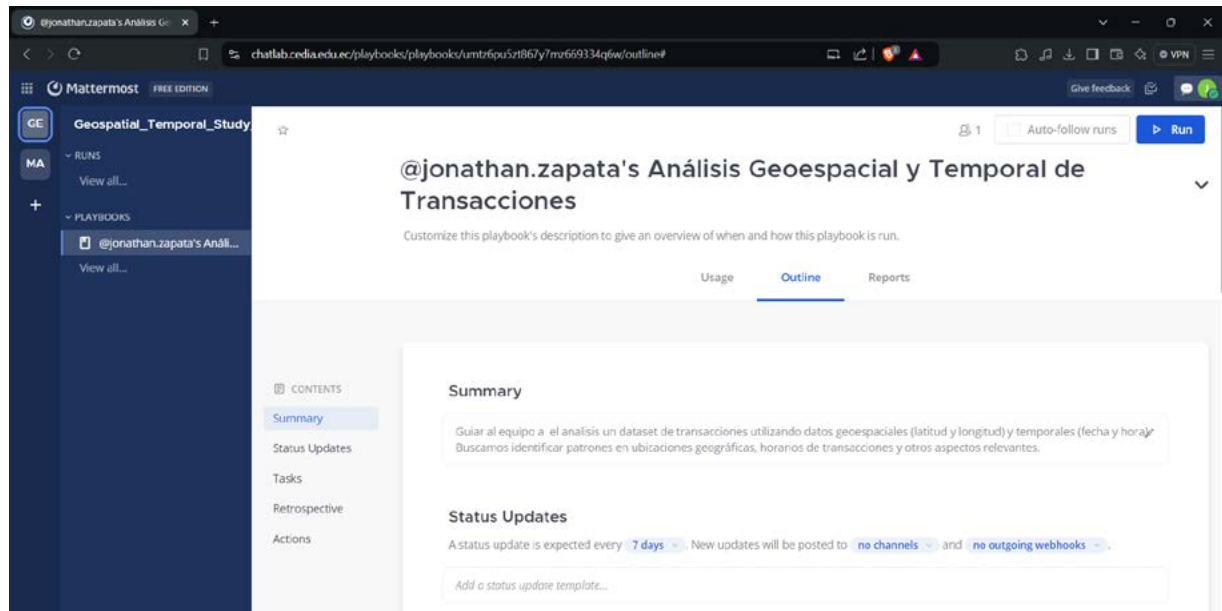


figura 70

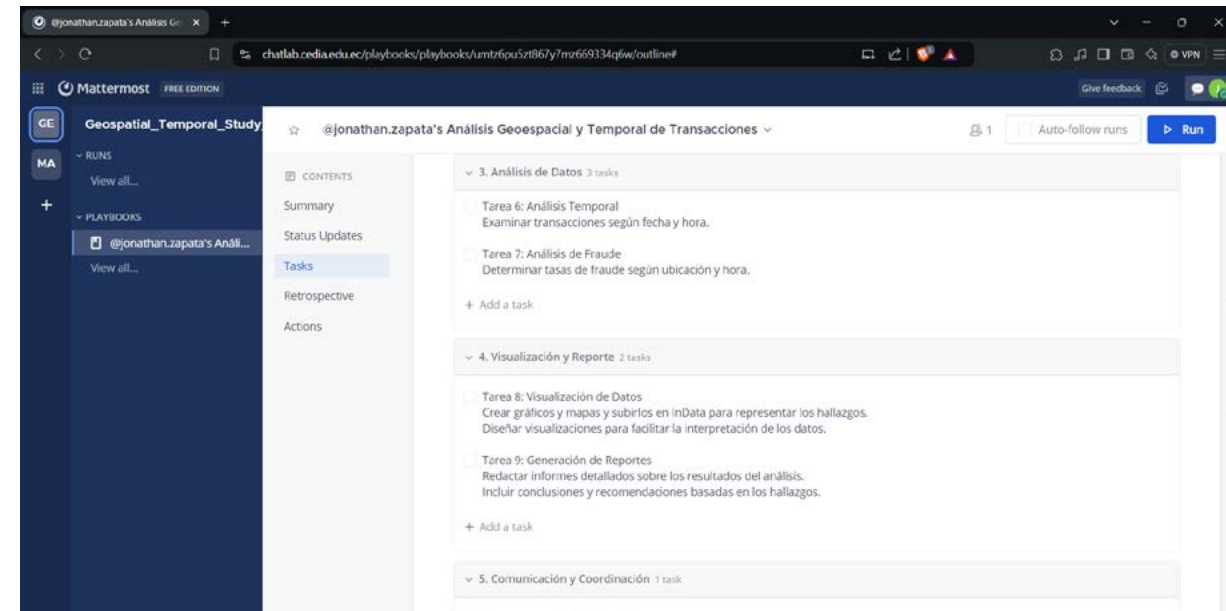


figura 72



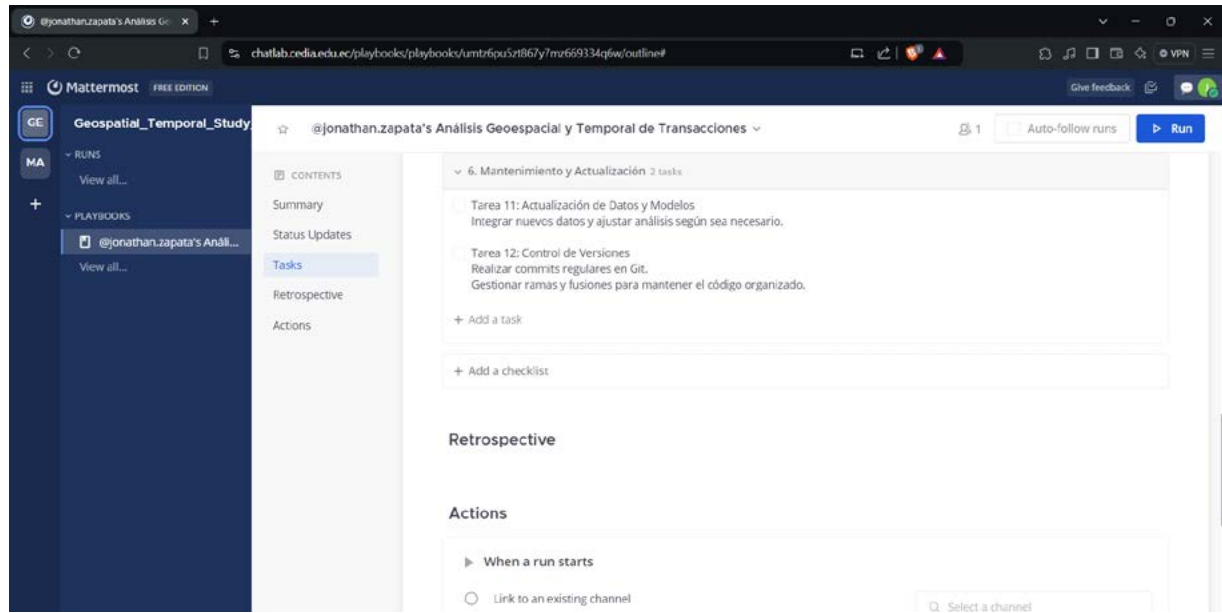


figura 73

Finalmente lo ejecutamos, con la configuración de canal de comunicación y participantes y asignación del rol del propietario

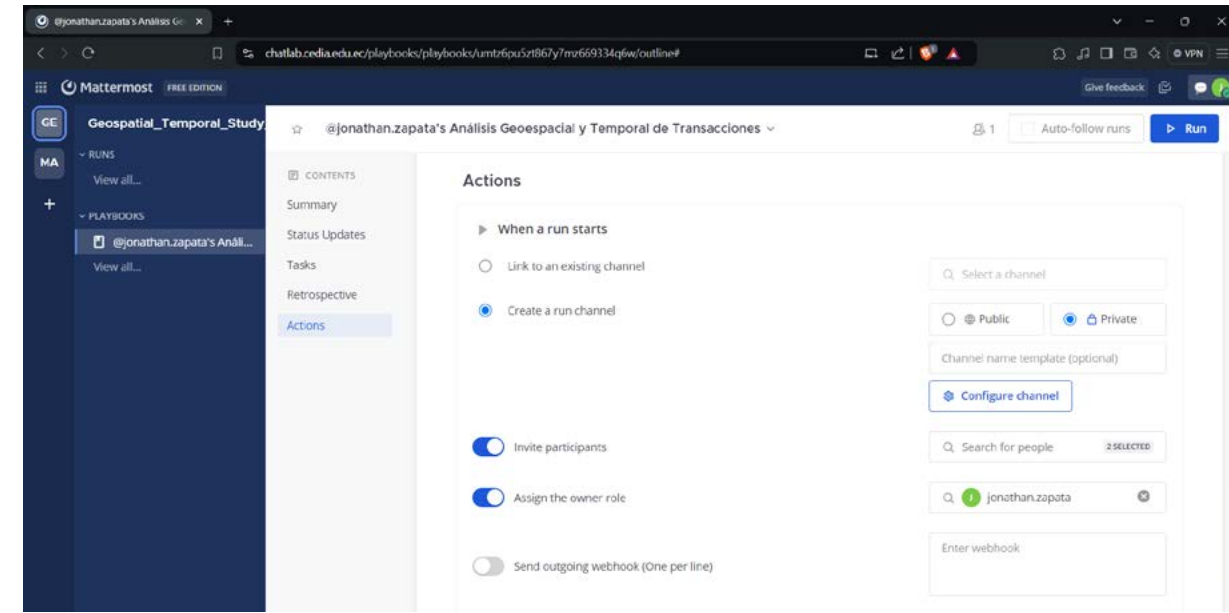


figura 74

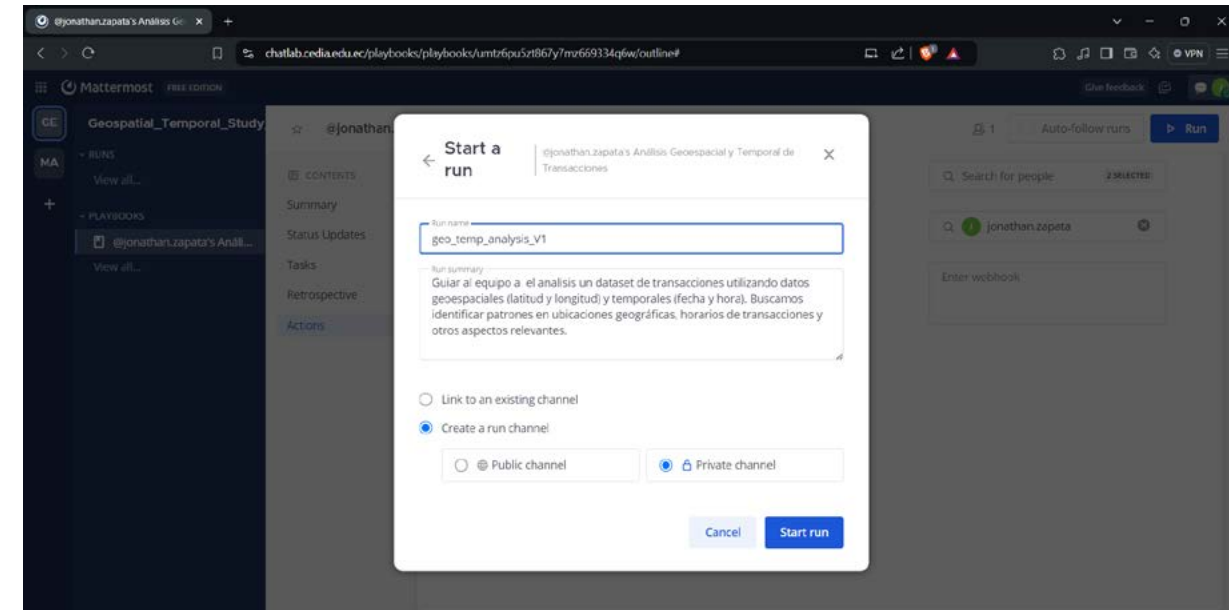


figura 75

Una vez que el playbook esté completo y ejecutado, podemos asignar los roles para cumplir con las tareas. Además, podemos detallar las acciones específicas para cada tarea y coordinar los comandos necesarios a través del chat .

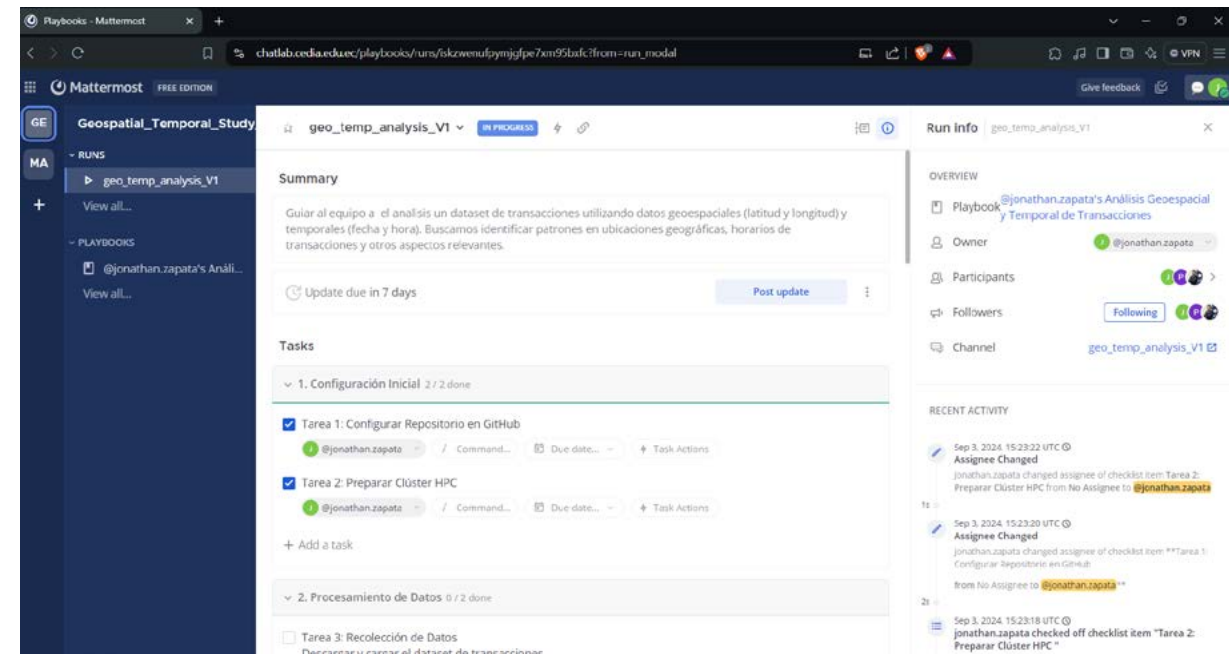


figura 76

Asi que tenemos un playbook bien elaborado que actúa como una guía integral que organiza y facilita el proceso de desarrollo de un modelo de aprendizaje automático, mejorando la eficiencia, la comunicación y la calidad del proyecto, del mismo modo podemos editar el playbook editato o agregando mas tareas.

En este caso configurado a una actualización en cada día, del mismo modo registrar las tareas ya hechas y realizar un seguimiento de estas, y con la designación de tareas a cada uno de los integrantes que fueron invitados al grupo.

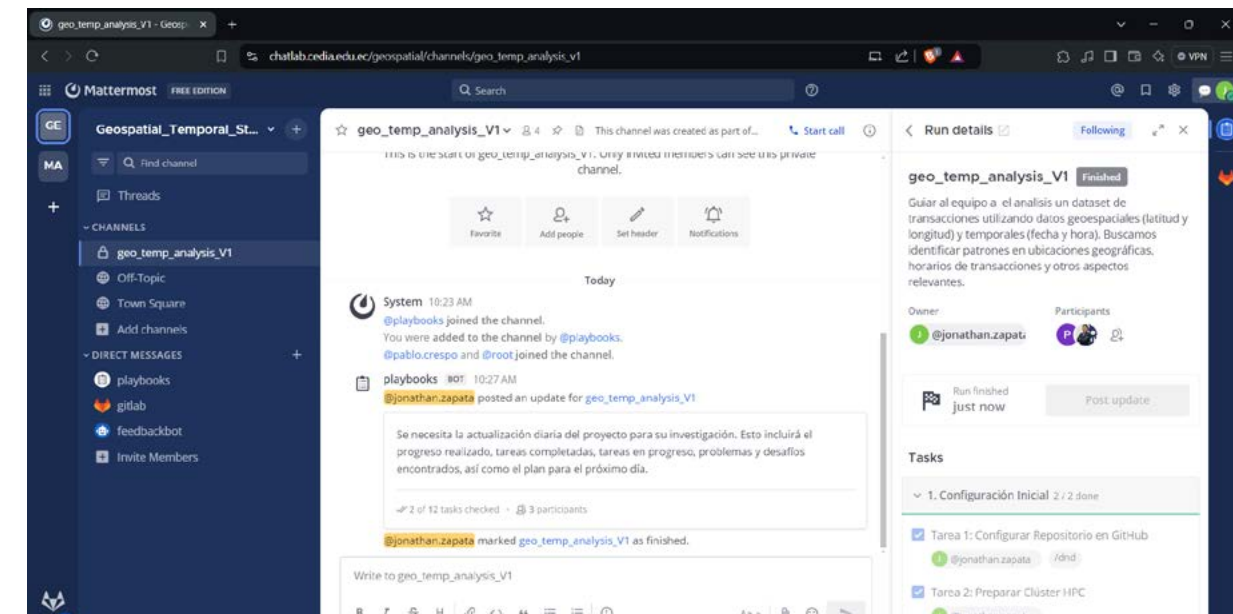


figura 77

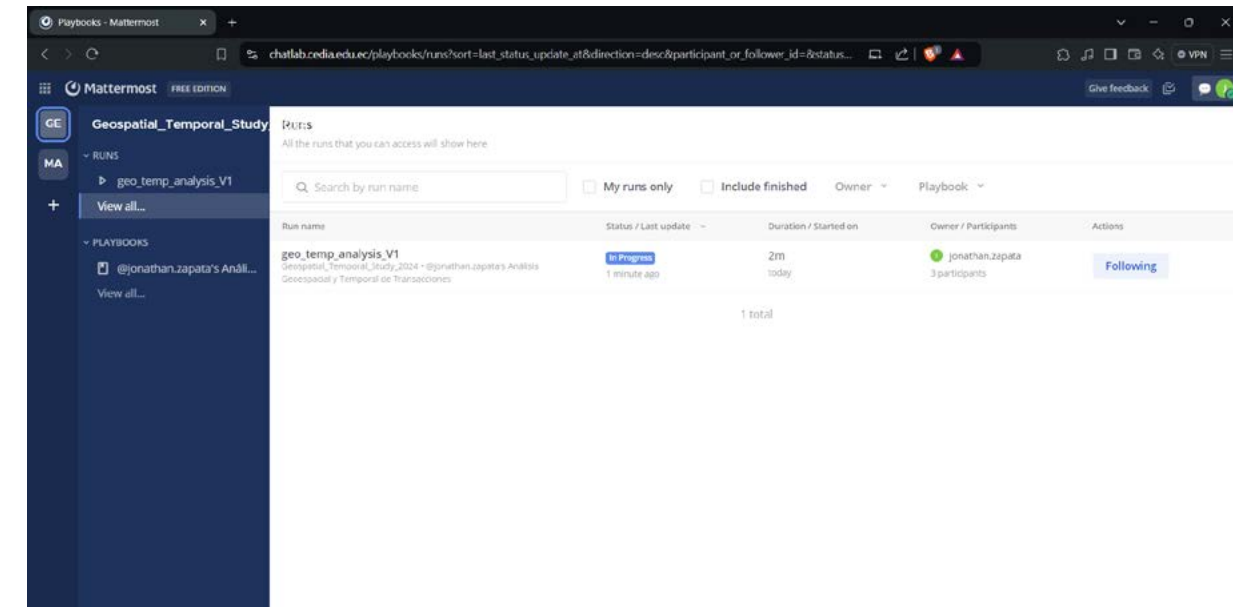


figura 78

Asimismo, podemos publicar una actualización del proyecto

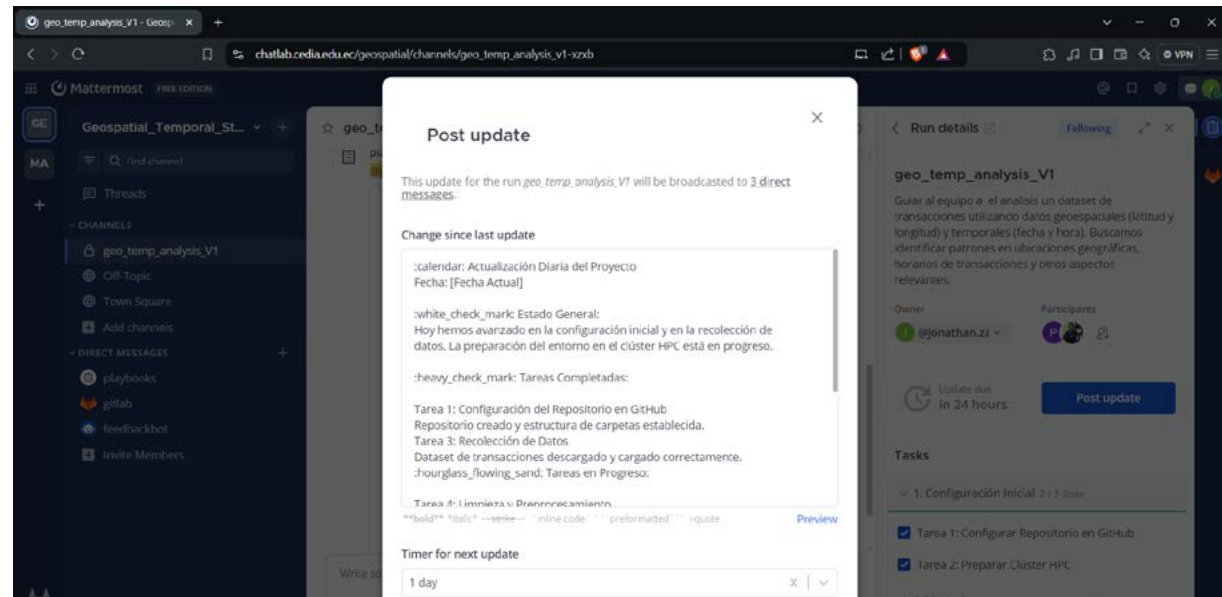


figura 79

Y ya finalizada las tareas podemos terminar el playbook.

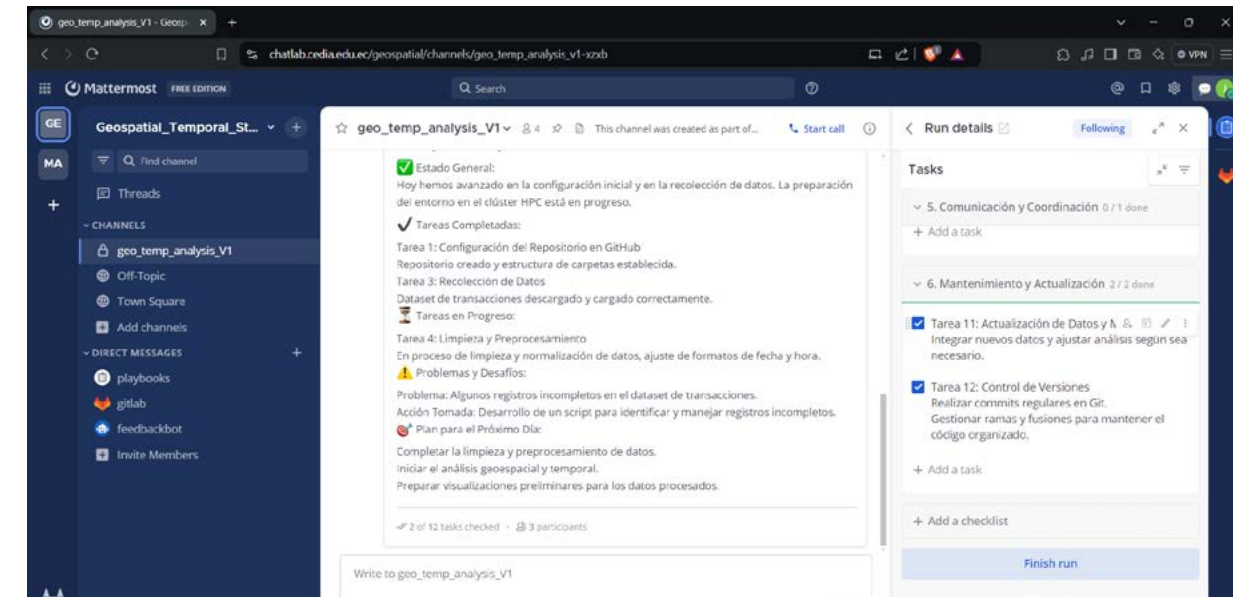


figura 81

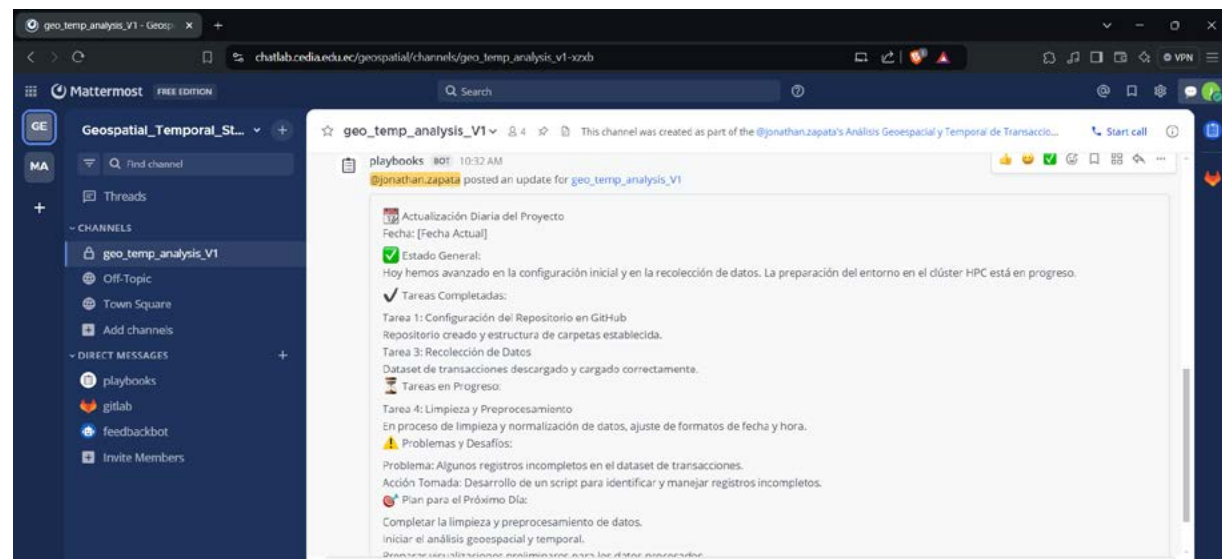


figura 80

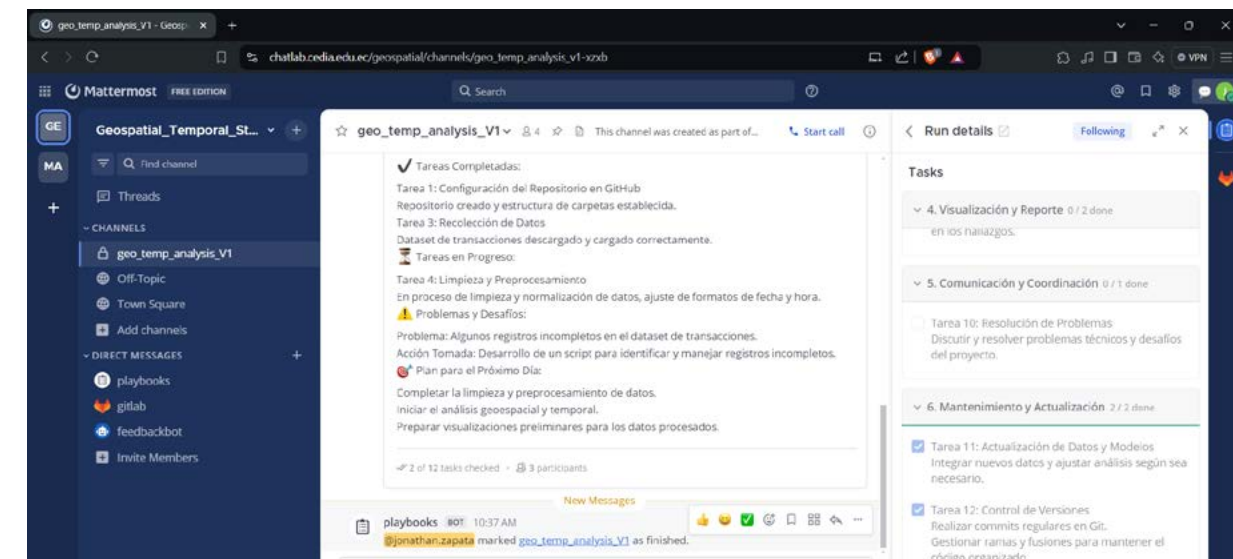


figura 82



# 1.6

## USO DE LA HERRAMIENTA FILESENDER



# filesender

Al utilizar la herramienta Filesender para enviar archivos de gran tamaño, designamos el archivo que queremos compartir—en este caso, nuestro dataset de aproximadamente 140 MB, así también los resultados de la investigación. Establecemos la fecha límite de expiración como el 13/09/2024 y procedemos a enviar el archivo al destinatario.

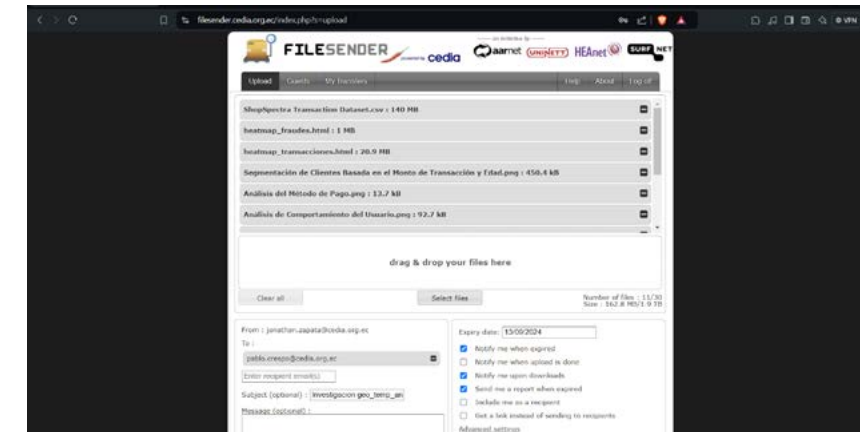


figura 83

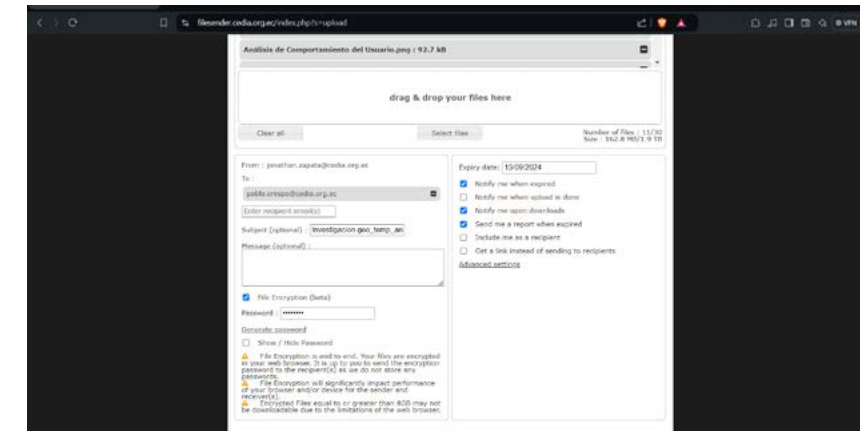


figura 84

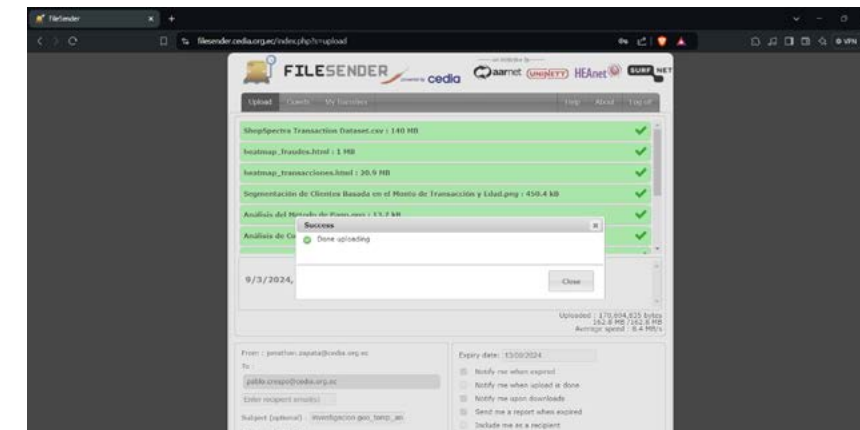


figura 85



**research**  
**lab**  
by cedia

cedia

*“Conectando ideas,  
transformando  
sociedades”*







[cedia.edu.ec](https://cedia.edu.ec)



**CUE**

Oficinas\_  
Gonzalo Cordero 2-122  
y J. Fajardo esquina.  
DIGITALS\_  
Miguel Moreno y Av. 10  
de Agosto

**UIO**

Av. 12 de Octubre y  
Lizardo García  
Edificio Alto Aragón  
Oficina 8A

**GYE**

Av. Del Bombero,  
Km 6.5 - Edificio  
La Vista de San Eduardo  
5to piso Ofics. 510 y 511

**PORTOVIEJO**

Av. Metropolitana  
Eloy Alfaro #2005  
y Av. Olímpica.  
Univ. San Gregorio

**MANTA**

Av. Circunvalación  
Vía a San Mateo  
ULEAM - EP