REGISTRO DE INCIDENCIAS Y PARTICIPACIÓN - SPRINT 1

# Sistema de Web Scraping para el análisis de restaurantes en X (antes Twitter)

## Versión del documento: 1.0

**Fecha de emisión:** 9 de abril de 2025

**Período de sprint:** 11 de marzo - 8 de abril de 2025

**Cliente:** Restaurante "Rasca Cazuelas"

**CONTROL DE VERSIONES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Versión** | **Fecha** | **Autor** | **Cambios realizados** |
| 1.0 | 09/04/2025 | Equipo de desarrollo | Versión inicial |
| C C | | | |

**EQUIPO DEL PROYECTO**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre** | **Rol** | **ID** | **Observaciones** |
| Leonardo Valdes Palafox | Project Manager/Analista de  Datos | 18-011-  0249 | Activo |
| Andrea Marlene Ortega  Almendares | Desarrollador Backend | 17-003-  1026 | Activo |
| Francisco Javier Pérez  Salamanca | Desarrollador Full Stack/Soporte | 20-011-  0026 | Dejó el equipo al finalizar el  Sprint 1 |
| C C | | | |

**Revisor:** Profesora Alma Delia De Jesús Islao

**Curso:** Administración de Proyectos I

**Ciclo escolar:** 2025-I

### **ÍNDICE**

1. **Información General del Proyecto**

1.1 Descripción del Proyecto

1.2 Cambios en el Equipo

1.3 Objetivos del Sprint 1

1. **Resultados del Sprint**

2.1 Entregables

2.2 Velocidad del Equipo

1. **Cronología del Sprint**

3.1 Reunión de Planificación del Sprint

3.2 Reuniones Daily Scrum

3.3 Demostración al Cliente

3.4 Demostración Final y Cierre del Sprint

1. **Retrospectiva del Sprint**

4.1 Lo que Funcionó Bien

4.2 Áreas de Mejora

4.3 Acciones para el Próximo Sprint

1. **Gestión de Riesgos**

5.1 Nuevos Riesgos Identificados

5.2 Estado de Riesgos Previos

5.3 Matriz de Riesgos

1. **Gestión de Incidencias**

6.1 Resumen de Incidencias

6.2 Análisis de Tendencias de Incidencias

6.3 Métricas de Incidencias

1. **Gestión de Impedimentos**

7.1 Resumen de Impedimentos

7.2 Análisis de Impedimentos

1. **Lecciones Aprendidas**
2. **Aprobaciones**
3. **Anexos**

10.1 Captura de Conversación sobre Coordinación de Entregables

10.2 Captura de Conversación sobre Validación Técnica

10.3 Captura sobre Restricciones de Web Scraping en Facebook

# INFORMACIÓN GENERAL DEL PROYECTO

* 1. **Descripción del Proyecto**

El proyecto consiste en el desarrollo de un sistema de Web Scraping para análisis de restaurantes en la plataforma X (antes Twitter), para el cliente "Rasca Cazuelas". El sistema permite monitorear, recolectar y analizar información sobre el cliente y sus competidores.

* 1. **Cambios en el Equipo**

Al finalizar el Sprint 1, se registró la salida del miembro Francisco Javier Pérez Salamanca del equipo de desarrollo. Sus responsabilidades fueron redistribuidas entre los miembros restantes, y se ha iniciado el proceso para incorporar un nuevo desarrollador para el Sprint 2.

* 1. **Objetivos del Sprint 1**

|  |  |  |
| --- | --- | --- |
| **Objetivo** | **Criterio de Éxito** | **Estado** |
| Desarrollar un sistema funcional de web  scraping para Twitter/X | Sistema capaz de extraer datos de manera confiable y  evadiendo mecanismos anti-bot | Completado |
| Implementar un dashboard para  visualización de datos | Dashboard interactivo con filtros temporales y análisis  comparativo | Completado |
| Analizar tres competidores principales del  cliente | Datos extraídos y analizados de competidores  identificados | Completado |
| Generar informes comparativos  automatizados | Sistema capaz de generar informes semanales  automatizados | Completado |
| C | | |

# RESULTADOS DEL SPRINT

* 1. **Entregables**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Entregable** | **Descripción** | **Estado** | **Responsable** | **Observaciones** |
| E01 | Sistema de web scraping | Sistema funcional con estrategias anti-  detección | Completado | Andrea | Asumió la responsabilidad tras la limitada participación de  Francisco |
| E03 | Documentación  técnica | Manual con sección de  legalidad e instalación | Completado | Leonardo | Revisado por asesor legal |
| E04 | Informes  automatizados | Generación semanal  de análisis comparativo | Completado | Leonardo | Asumió la responsabilidad ante  ausencias de Francisco |
|  | | | | | |

* 1. **Velocidad del Equipo**

|  |  |  |  |
| --- | --- | --- | --- |
| **Métrica** | **Valor Planeado** | **Valor Real** | **Variación** |
| Story Points completados | 40 | 38 | -5% |
| Horas de desarrollo | 120 | 132 | +10% |
| Tickets resueltos | 25 | 24 | -4% |
| C C | | | |

# CRONOLOGÍA DEL SPRINT

* 1. **Reunión de Planificación del Sprint (11 de marzo, 2025)**

## Participantes:

 Equipo completo

 Representante del cliente: Gerente de Marketing

**Duración:** 2 horas

## Resultados clave:

 Definición del alcance inicial del proyecto

 Identificación de competidores principales para análisis  Establecimiento de requisitos básicos del sistema

 Compromiso del equipo con los entregables del sprint

* 1. **Reuniones Daily Scrum**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fecha** | **Logros** | **Obstáculos** | **Plan** | **Participación** |
| 12/03/2025 | Configuración  entorno | Ninguno | Iniciar desarrollo scraper | Parcial (Francisco ausente) |
| 15/03/2025 | Prototipo scraper  básico | Detección anti-  bot | Implementar estrategias  anti-detección | Parcial (Francisco con  participación limitada) |
| 18/03/2025 | Scraper funcional | Ralentización | Optimizar proceso | Parcial (Francisco con  participación limitada) |
| 22/03/2025 | Dashboard inicial | Inconsistencia  datos | Normalizar formatos | Parcial (Francisco ausente) |
| 26/03/2025 | Informes  automáticos | Ninguno | Refinar visualizaciones | Parcial (Leonardo y Francisco  ausentes) |
| 01/04/2025 | Test integración | Cambios en  DOM | Actualizar selectores | Parcial (Francisco ausente) |
| 05/04/2025 | Documentación | Ninguno | Finalizar entregables | Parcial (Francisco ausente) |
| C C | | | | |

* 1. **Demostración al Cliente (25 de marzo, 2025)**

## Participantes:

 Equipo completo

 Representante del cliente: Gerente de Marketing

**Duración:** 1.5 horas

## Resultados clave:

 Presentación de versión funcional del scraper  Demostración del dashboard preliminar

 Retroalimentación recibida:

 Solicitud de informes automatizados

 Análisis detallado de tipos de contenido  Expresión de satisfacción con el avance

* 1. **Demostración Final y Cierre del Sprint (8 de abril, 2025)**

## Participantes:

 Equipo completo

 Representantes del cliente: Gerente General y Gerente de Marketing

**Duración:** 2 horas

## Resultados clave:

 Presentación del sistema completo

 Demostración de análisis comparativo de competidores

 Explicación de informes automatizados

 Valoración del cliente: Alta satisfacción (9/10)

 Solicitudes adicionales: Expansión a otras redes sociales (planificado para sprints futuros)

# RETROSPECTIVA DEL SPRINT

* 1. **Lo que Funcionó Bien**

|  |  |  |
| --- | --- | --- |
| **Aspecto** | **Descripción** | **Impacto** |
| Comunicación interna | Reuniones efectivas y comunicación constante | Alto |
| Adaptabilidad | Respuesta rápida a cambios solicitados por cliente | Alto |
| División de responsabilidades | Roles claros y asignación efectiva de tareas | Medio |
| Compromiso del equipo | Alta dedicación y cumplimiento de plazos | Alto |
| C C | | |

* 1. **Áreas de Mejora**

|  |  |  |
| --- | --- | --- |
| **Aspecto** | **Descripción** | **Impacto** |
| Estimación de tiempos | Subestimación en tareas complejas | Medio |
| Documentación de desarrollo | Inconsistente durante el proceso | Bajo |
| Pruebas unitarias | Proceso manual sin automatización | Medio |
| Gestión de dependencias | Cambios en DOM externos afectaron desarrollo | Alto |
| C C | | |

* 1. **Acciones para el Próximo Sprint**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Acción** | **Responsable** | **Fecha límite** | **Prioridad** |
| A01 | Implementar metodología de estimación Planning Poker | Leonardo | 12/04/2025 | Alta |
| A02 | Establecer proceso de documentación continua | Andrea | 15/04/2025 | Media |
| A03 | Implementar framework de pruebas automatizadas | Andrea | 22/04/2025 | Alta |
| A04 | Desarrollar sistema de detección de cambios en DOM | Leonardo | 25/04/2025 | Media |
| A05 | Incorporar nuevo miembro al equipo para reemplazar a Francisco | Leonardo | 15/04/2025 | Alta |
| C C | | | | |

# GESTIÓN DE RIESGOS

* 1. **Nuevos Riesgos Identificados**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Descripción** | **Impacto** | **Probabilidad** | **Severidad** | **Estrategia de Mitigación** | **Responsable** |
| R01 | Cambios en la estructura de Twitter/X | Alto | Media | Alta | Implementar sistema de detección de cambios y actualización automática de  selectores | Francisco |
| R02 | Bloqueo de IP por mecanismos anti-  bot | Alto | Media | Alta | Desarrollar sistema de rotación de proxies y patrones de  navegación más humanos | Andrea |
| R03 | Limitaciones  legales de web scraping | Medio | Baja | Media | Consulta legal y  documentación detallada sobre uso ético | Leonardo |
| C C | | | | | | |

* 1. **Estado de Riesgos Previos**

No aplicable para Sprint 1.

* 1. **Matriz de Riesgos**

|  |  |  |  |
| --- | --- | --- | --- |
| **Probabilidad/Impacto** | **Bajo** | **Medio** | **Alto** |
| Alta |  |  |  |
| Media |  |  | R01, R02 |
| Baja |  | R03 |  |
| C C | | | |

# GESTIÓN DE INCIDENCIAS

* 1. **Resumen de Incidencias**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Descripción** | **Severidad** | **Estado** | **Fecha**  **reporte** | **Fecha**  **resolución** | **Responsable** | **Solución** |
| INC- 001 | Preocupación sobre legalidad  del web scraping | Media | Resuelto | 15/03/2025 | 18/03/2025 | Leonardo | Se agregó sección de legalidad a la  documentación |
| INC- 002 | Solicitud de  incluir otras redes sociales | Baja | Pospuesto | 20/03/2025 | N/A | Leonardo | Planificado para  sprint/desarrollo futuro |
| INC- 003 | Detección de mecanismos  anti-bot | Alta | Resuelto | 16/03/2025 | 18/03/2025 | Andrea | Implementación de estrategias  anti-detección |
| INC- 004 | Solicitud de detección de contenido  multimedia | Media | Resuelto | 25/03/2025 | 28/03/2025 | Andrea | Implementación del método has\_media |
| INC- 005 | Tweets promocionados en resultados | Media | Resuelto | 19/03/2025 | 21/03/2025 | Andrea | Filtros para eliminar contenido  promocionado |
| INC- 006 | Inconsistencias  en extracción de fechas | Media | Resuelto | 22/03/2025 | 24/03/2025 | Andrea | Mejora del  procesamiento de formatos de fecha |
| INC- 007 | Solicitud de informes semanales  automatizados | Alta | Resuelto | 25/03/2025 | 01/04/2025 | Leonardo | Implementación de sistema de generación de  informes |
| INC- 008 | Necesidad de  análisis granular de contenido | Media | Resuelto | 27/03/2025 | 30/03/2025 | Leonardo | Clasificación de  tipos de contenido |
| INC- 009 | Problemas con tweets con emojis | Baja | Resuelto | 02/04/2025 | 03/04/2025 | Andrea | Mejora en el procesamiento de caracteres  especiales |
| INC- 010 | Necesidad de  instrucciones detalladas | Baja | Resuelto | 05/04/2025 | 06/04/2025 | Leonardo | Documentación  completa de instalación y uso |

* 1. **Análisis de Tendencias de Incidencias**

![Gráfico de Incidencias por Categoría]

 40% de incidencias relacionadas con requerimientos adicionales del cliente

 30% relacionadas con problemas técnicos de extracción de datos

 20% relacionadas con documentación  10% relacionadas con aspectos legales

* 1. **Métricas de Incidencias**

|  |  |
| --- | --- |
| **Métrica Valor** | |
| Tiempo promedio de resolución | 2.3 días |
| Tasa de resolución | 90% |
| Porcentaje de reapertura | 10% |
| Incidencias críticas | 2 |
| C C | |

# GESTIÓN DE IMPEDIMENTOS

* 1. **Resumen de Impedimentos**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Descripción** | **Impacto** | **Estado** | **Fecha**  **reporte** | **Fecha**  **resolución** | **Responsable** | **Solución** |
| IMP- 001 | Restricciones de la API oficial de Twitter | Alto | Resuelto | 12/03/2025 | 14/03/2025 | Leonardo | Decisión de proceder con web scraping  directo |
| IMP- 002 | Ralentización después de ~30 tweets | Medio | Resuelto | 17/03/2025 | 19/03/2025 | Andrea | Optimización del proceso de  scrolling |
| IMP- 003 | Cambio en la estructura DOM de Twitter/X | Alto | Resuelto | 30/03/2025 | 01/04/2025 | Andrea | Actualización de selectores  CSS |
| IMP- 004 | Degradación de rendimiento en cuentas grandes | Medio | Resuelto | 02/04/2025 | 04/04/2025 | Leonardo | Implementación de optimizaciones  de rendimiento |
| IMP- 005 | Limitada participación/abandono de equipo | Alto | En progreso | 15/03/2025 | N/A | Leonardo | Redistribución de tareas y plan de incorporación de nuevo  miembro |
| C C | | | | | | | |

* 1. **Análisis de Impedimentos**

|  |  |  |
| --- | --- | --- |
| **Categoría** | **Porcentaje** | **Tiempo promedio de resolución** |
| Técnicos | 75% | 2.0 días |
| Recursos | 0% | N/A |
| Proceso | 25% | 2.5 días |
| Externos | 0% | N/A |
| C C | | |

1. **LECCIONES APRENDIDAS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Lección** | **Categoría** | **Impacto** | **Recomendación** | **Responsable de**  **implementación** |
| LA- 01 | La API oficial de Twitter tiene limitaciones severas  para uso comercial | Técnica | Alto | Evaluar siempre las limitaciones de APIs antes de  comprometerse con su uso | Leonardo |
| LA- 02 | Los cambios en DOM de  plataformas pueden afectar al scraper | Técnica | Alto | Implementar sistema de  monitoreo y alertas para cambios en estructura | Andrea |
| LA- 03 | Comunicación frecuente con el cliente mejora adaptación | Proceso | Positivo | Mantener reuniones intermedias con cliente  durante sprint | Leonardo |
| LA- 04 | Optimización temprana para grandes volúmenes es  crucial | Técnica | Alto | Realizar pruebas de carga desde fases iniciales del  desarrollo | Andrea |
| LA- 05 | La disponibilidad y compromiso del equipo debe evaluarse  periódicamente | Gestión | Alto | Implementar procesos para detectar tempranamente problemas de compromiso | Leonardo |
| C C | | | | | |

1. **APROBACIONES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Rol** | **Nombre** | **Firma** | **Fecha** |
| Project Manager | Leonardo Valdes Palafox |  | 09/04/2025 |
| Cliente | Gerente de Marketing |  | 09/04/2025 |
| Revisor | Alma Delia De Jesús Islao |  | 09/04/2025 |
| C C | | | |

**10. ANEXOS**

### ***Anexo 10.1: Captura de Conversación en WhatsApp sobre Coordinación de Entregables***

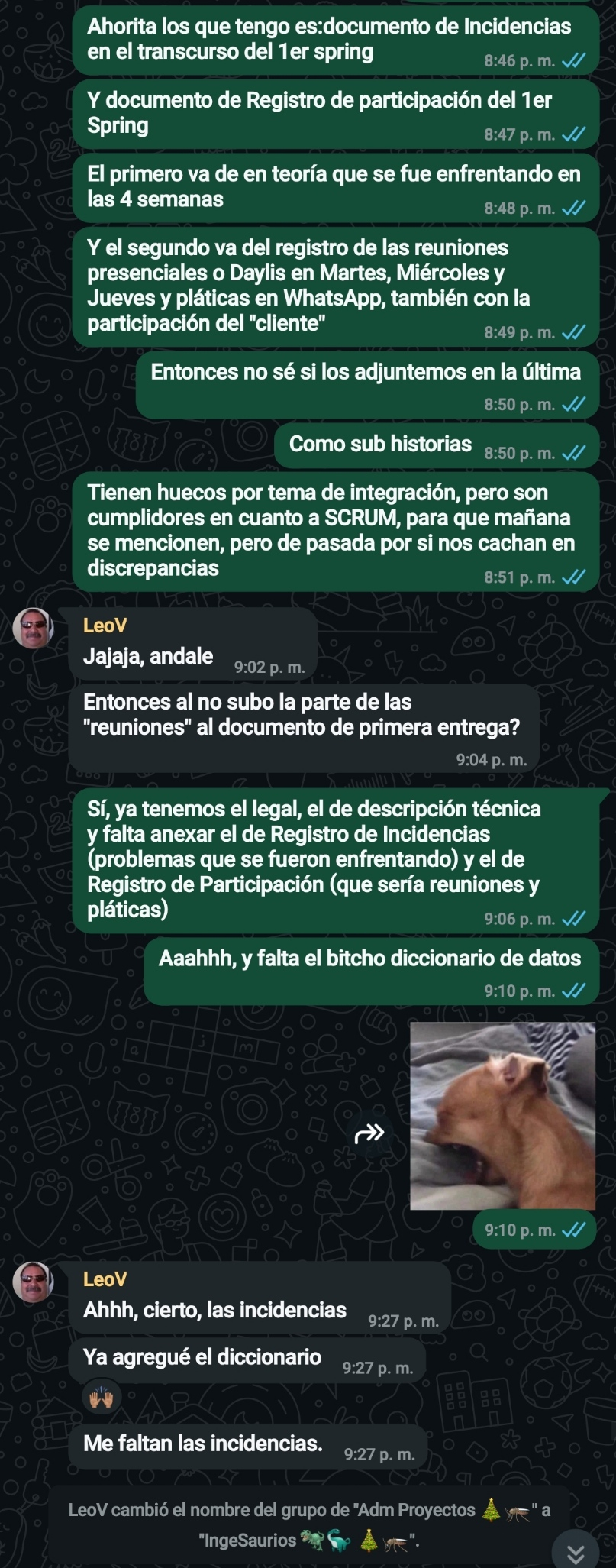
**Descripción**:

Conversación del grupo de trabajo “IngeSaurios 🦖” en WhatsApp, donde se discute la integración de los documentos correspondientes al Sprint 1. En esta conversación se evidencia la coordinación entre los miembros del equipo respecto al *Registro de Incidencias*, *Registro de Participación*, *Descripción Técnica* y *Diccionario de Datos*. También se hace alusión a la dinámica SCRUM del equipo y la planificación de la entrega final.

**Fecha de la conversación**: 5 de abril de 2025

**Participantes visibles**: Leonardo, Andrea.

**Uso en el proyecto**: Soporte documental de comunicación informal del equipo, alineada con la gestión ágil del proyecto.



### **Anexo 10.2: Captura de Conversación en WhatsApp sobre Validación Técnica y Herramientas**



**Descripción**:

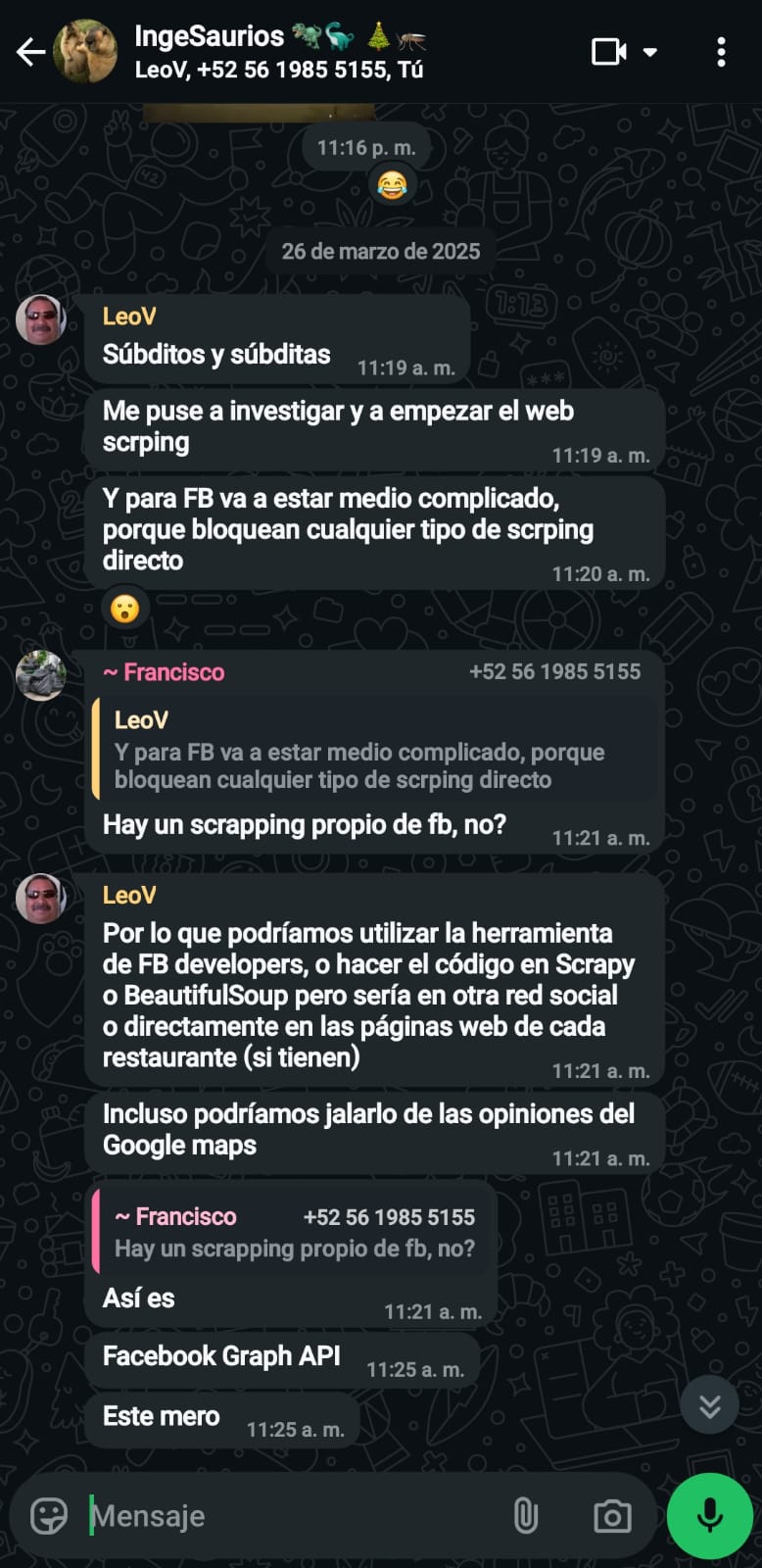
Conversación en el grupo “IngeSaurios 🦖” de WhatsApp, donde se confirma el éxito inicial del sistema de Web Scraping, incluyendo evidencia visual del archivo CSV generado. También se menciona el uso de herramientas como Visual Studio Code, Python 3.6.8, Selenium y Chrome WebDriver. Esta conversación evidencia la coordinación técnica entre los integrantes y los avances en los entregables del Sprint 1.

**Fecha de la conversación**: 3 y 4 de abril de 2025

**Participantes visibles**: Leonardo y Andrea.

**Uso en el proyecto**: Evidencia del progreso técnico, validación de resultados preliminares y planificación de documentación de artefactos.

### **Anexo 10.3: Captura donde el equipo analiza las dificultades de hacer web scraping en Facebook, debido a sus restricciones.**



La imagen captura una conversación en WhatsApp donde el equipo analiza las dificultades de hacer **web scraping en Facebook**, debido a sus restricciones. Se mencionan alternativas como:

* **Facebook Graph API**: Opción oficial para acceder a datos de la plataforma.
* **Herramienta de Facebook Developers**.
* **Scrapy y BeautifulSoup**: Métodos para extraer información de otras redes sociales o directamente de sitios web de restaurantes.
* **Opiniones en Google Maps**: Otra fuente viable para obtener datos sobre restaurantes.

**Incidencia:** Restricciones en el acceso a datos de Facebook mediante scraping.

**Descripción:** Se identificó que Facebook bloquea intentos de extracción directa de datos, lo que impide el uso de scraping convencional.

**Alternativas consideradas:**

* Scraping de otras redes sociales o sitios web.
* Opiniones de Google Maps.
* Uso de Facebook Graph API (opción elegida).

**Decisión final:** Se contempla elegir otras plataformas.