

## Ironhack Payments

El objetivo del proyecto es analizar el sistema de pago Ironhack Payments utilizando datos recopilados en dos archivos CSV. Nuestro objetivo principal es mejorar nuestros servicios identificando patrones y tendencias basados en métricas clave.

## Análisis Exploratorio de Datos

### Definición de Cohortes

Segmentar a los usuarios en grupos según su primer mes de negociación es clave para comprender cómo cambia el comportamiento de los clientes con el tiempo.

En nuestro caso lo haremos dividiendo a los usuarios por el mes de su primera transacción. Este método permite una evaluación más precisa y contextualizada de las métricas seleccionadas.

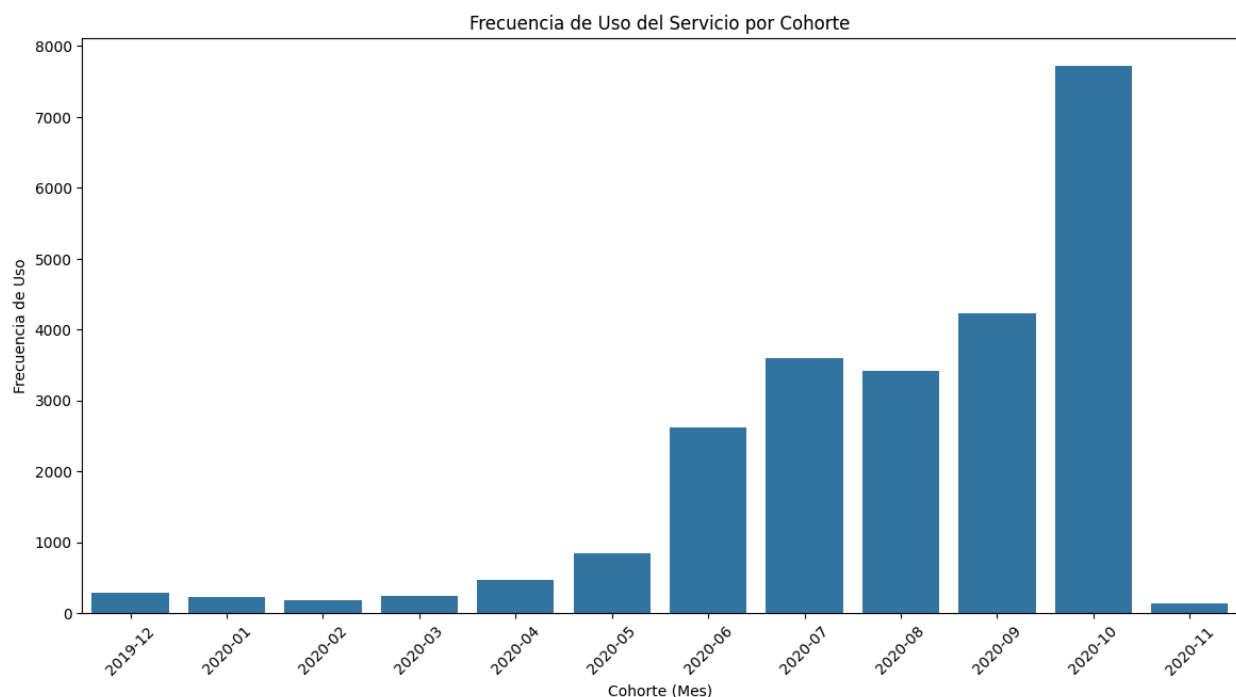
### Objetivos de la Etapa Inicial

En esta etapa inicial, el objetivo ha sido estudiar las cohortes según las siguientes métricas:

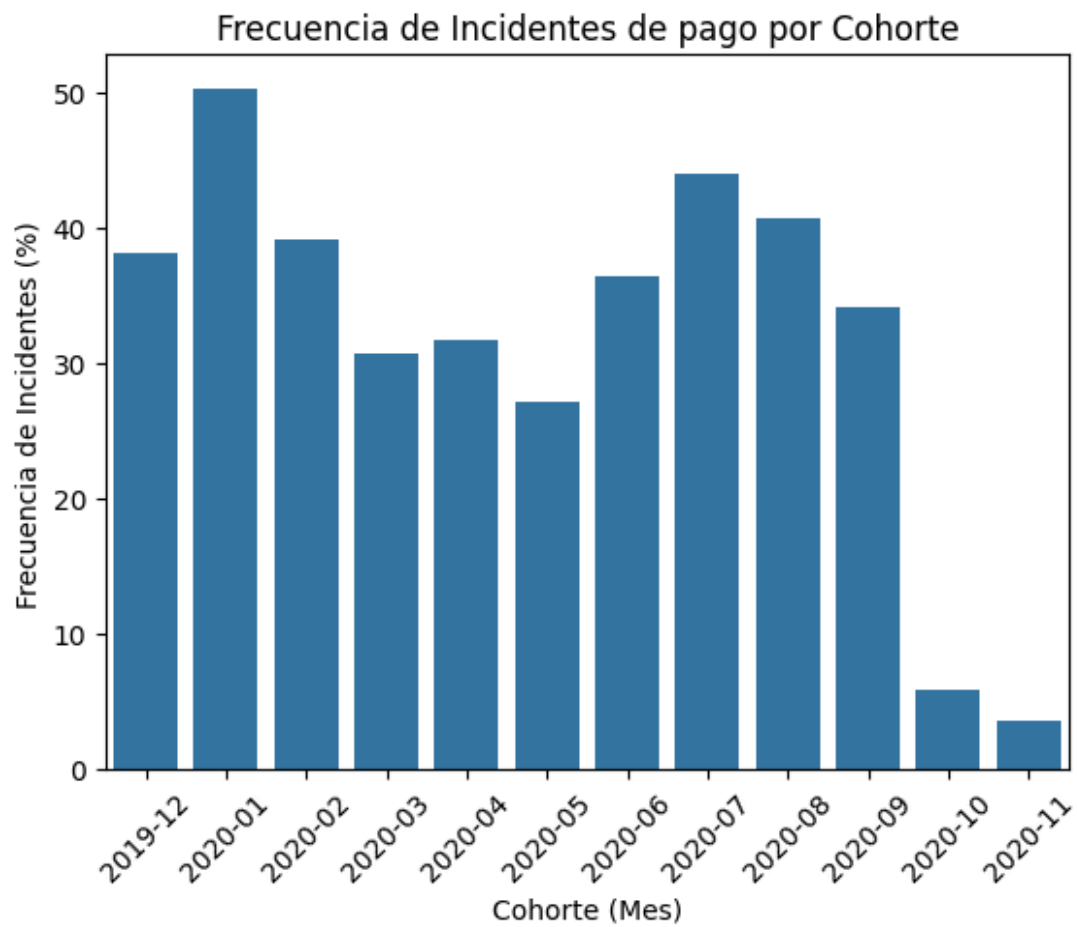
- Frecuencia de uso
- Frecuencia de incidentes
- Ingresos generados
- Tasa de incidentes vs. frecuencia de uso

### Frecuencia de uso

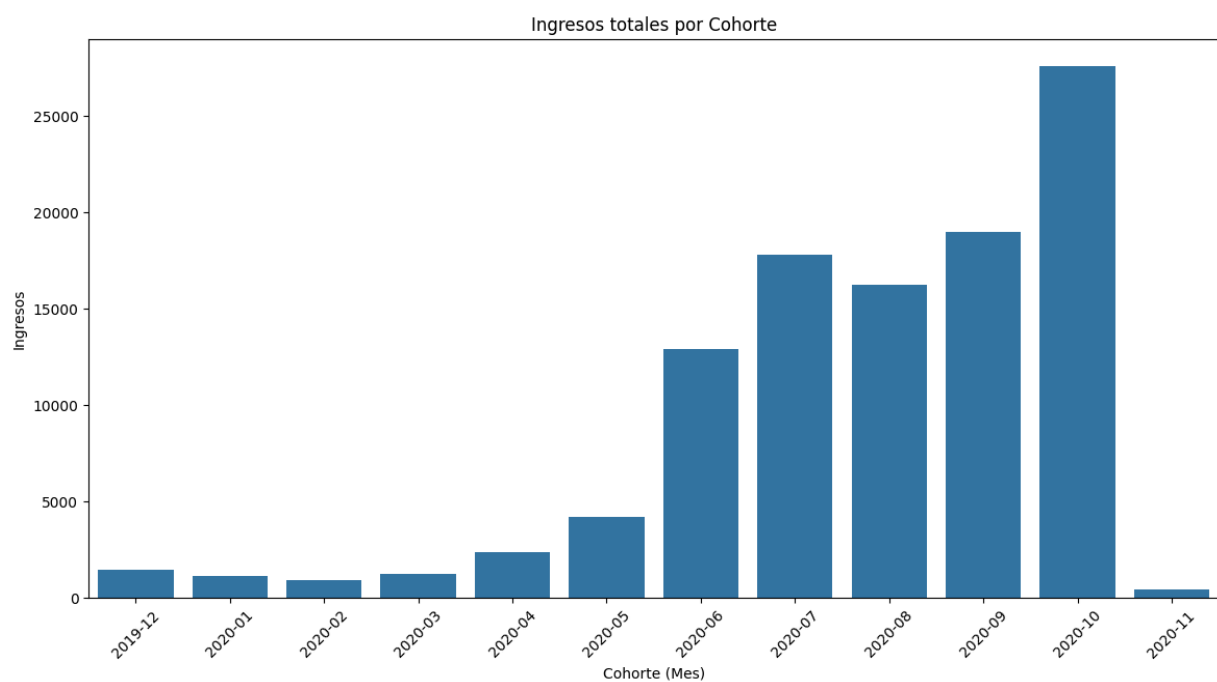
Desde el inicio del servicio se observa un aumento de la frecuencia del uso progresiva, hasta la cohorte del 10-2020 que muestra una frecuencia mucho más elevada que el resto. Y luego la del 11-2020 vuelve a niveles muy bajos.



## Frecuencia de incidentes

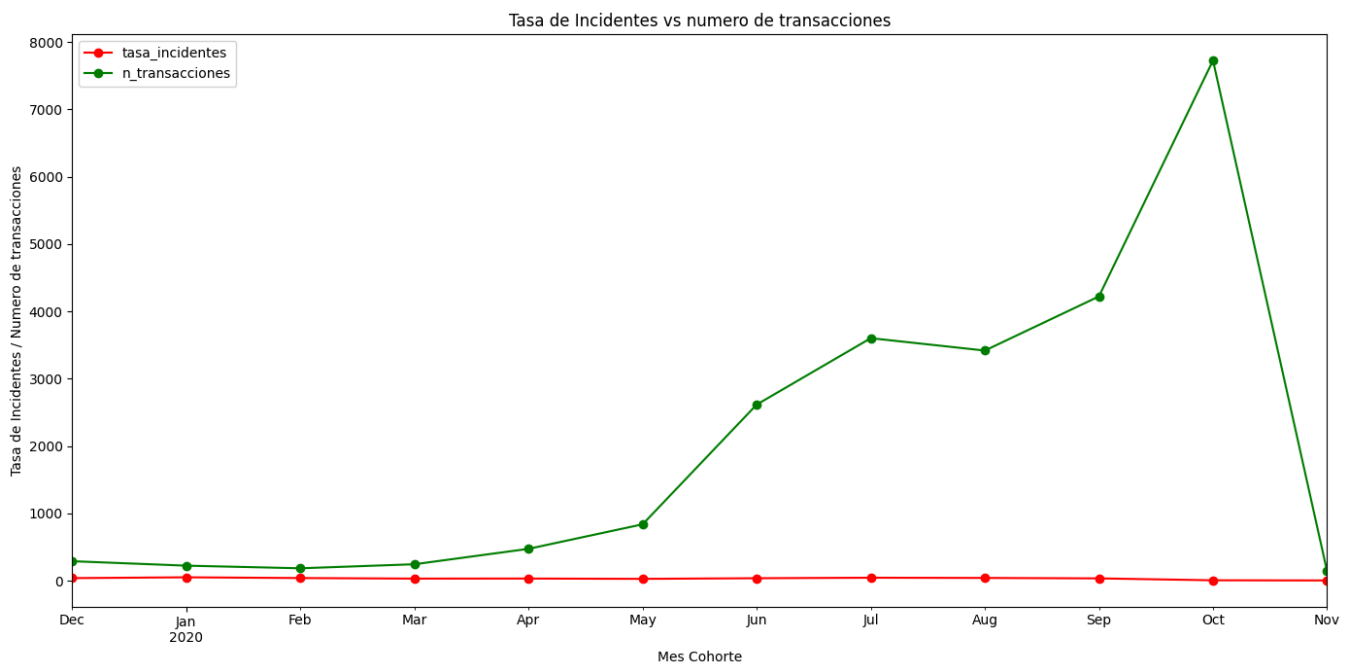


## Ingresos generados



### Tasa de incidentes vs. frecuencia de uso

No se observa una relación entre el número de transacciones y la tasa de incidentes, lo cual sugiere que el sistema es capaz de manejar eficientemente el volumen generado. Esto indica que el sistema no se ve afectado por el aumento en el número de transacciones en términos de incidencias, lo que subraya su capacidad para manejar eficazmente cualquier volumen de transacciones que se genere.



# Limpieza de Datos

## Detección de Valores Problemáticos

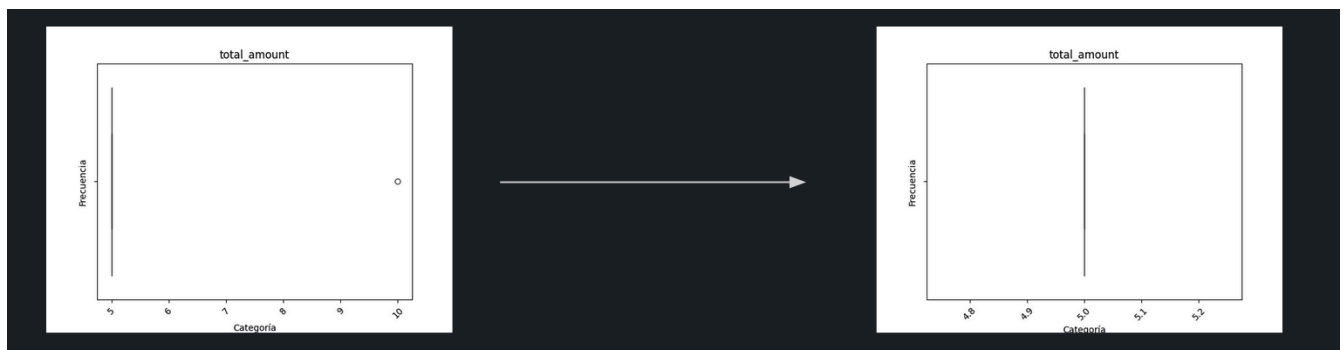
A partir del análisis inicial, se detectaron una serie de valores que afectaban negativamente las métricas observadas. Por ello, se realizó una limpieza de datos.

Para la limpieza y análisis de datos se utilizaron herramientas avanzadas como Python con Pandas y las bibliotecas NumPy.

Estas herramientas facilitan la detección y corrección de valores atípicos y la gestión de datos faltantes. Además, se utilizaron métodos estadísticos para la extracción de datos y el modelado predictivo.

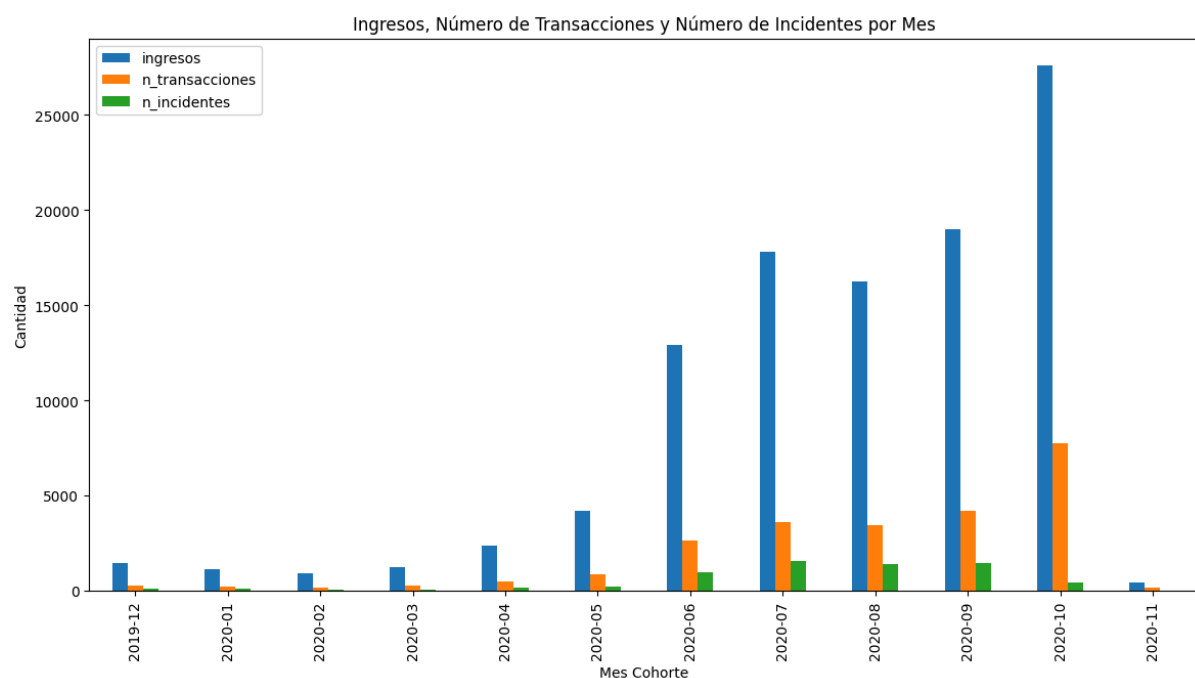
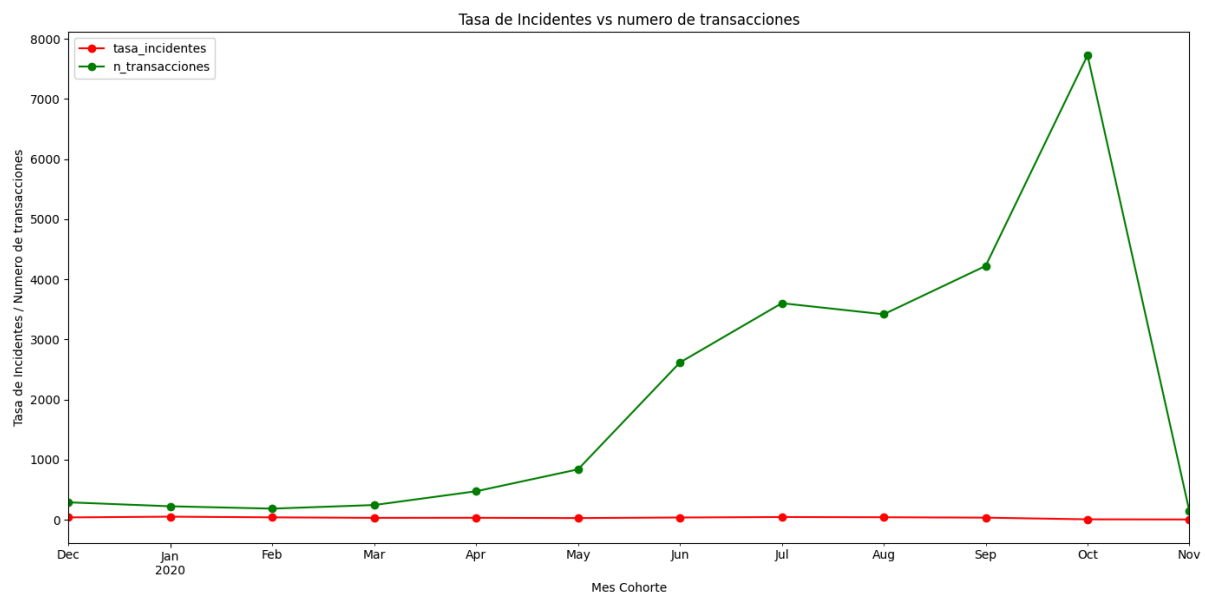
## Pasos de la Limpieza de Datos

1. Detección de Valores Outlier:
  - Identificamos y eliminamos valores atípicos que distorsionan las estadísticas, asegurando valores más representativos.
2. Manejo de Valores Missing:
  - Detectamos valores faltantes y los sustituimos adecuadamente para mantener la integridad de los datos.



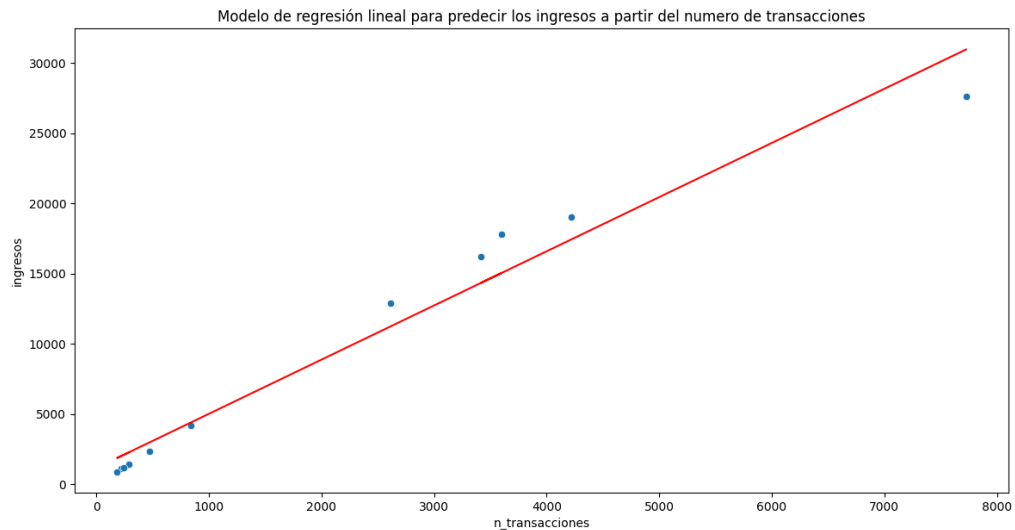
## Conclusiones

1. Robustez del sistema:
  - a. No se observa una relación entre el número de transacciones y la tasa de incidentes, lo que indica que el sistema es capaz de manejar eficientemente el volumen generado.
  - b. Aumento de Transacciones e Ingresos:
  - c. A partir de la cohorte de abril de 2020, se observa una subida significativa en el número de ingresos y transacciones.
2. Cohorte de Octubre 2020:
  - a. En la cohorte de octubre de 2020, se registra un número récord de transacciones y una cantidad absoluta de incidentes muy baja. Es necesario investigar más a fondo las causas de este mes excepcionalmente positivo.
3. Relación entre Ingresos y Transacciones:
  - a. El modelo de regresión lineal muestra que un aumento en el número de transacciones generará un aumento proporcional en los ingresos.



#### 4. Relación entre Ingresos y Transacciones:

- a. El modelo de regresión lineal muestra que un aumento en el número de transacciones generará un aumento proporcional en los



#### Tendencia Positiva

- La línea de regresión muestra una tendencia positiva, lo que indica que a medida que aumenta el número de transacciones, también aumentan los ingresos. Esto sugiere una relación directa entre estas dos variables.

#### Ajuste del Modelo

- La mayoría de los puntos de datos se encuentran cerca de la línea de regresión, lo que indica que el modelo tiene un buen ajuste. Sin embargo, algunos puntos se desvían de la línea, lo que puede sugerir la presencia de posibles outliers o variabilidad en los datos que no es capturada completamente por el modelo lineal.