**INSTITUTO POLITÉCNICO NACIONAL**

UNIDAD PROFESIONAL INTERDISCIPLINARIA   
DE INGENIERÍA Y CIENCIAS SOCIALES Y ADMINISTRATIVAS

SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

MAESTRÍA EN CIENCIAS EN INFORMÁTICA

**Diseño de proyecto BI**

**Proyecto ABIRID. Análisis con Business Intelligence del Robo de IDentidad**

Profesor: M. en C. Rafael Ibáñez Castañeda

Alumno: Adolfo Bravo Hernández

Ciudad de México, abril de 2018

Resumen del proyecto

# El ámbito: ¿Qué deberemos entregar?

Se entregará un proyecto de inteligencia de negocios, en donde se obtengan y aprovechen datos que tengan una fuerte correlación con el Robo de Identidad en México. Con lo que se podrán medir su comportamiento.

# El plazo: ¿Cuándo lo tendremos que entregar?

La fecha límite de entrega del proyecto es el 28 de mayo de 2018.

# El costo: ¿Cuánto costará? Y Los recursos: ¿Quién lo hará?

El proyecto no tendrá costo, pues se realiza sin utilizar recursos adicionales a los propios.

Haciendo un ejercicio para el proyecto hasta entes del mantenimiento, o cambios en el alcance se obtuvo un costo de $98, 703.70

|  |  |  |
| --- | --- | --- |
| COSTO POR USO DE HARDWARE  Laps de consultores + 2 Micro para la oficina arrendadas  Micro : Intel Core i5 3ra G. Pantalla : 14 pulgadas Stock : 5  Fuente: http://www.alquilerdepc.com/alquiler\_de\_pc\_precios.php |  | $1,800.00 |
| LICENCIA POWER BI 8.90 euros por mes |  | $204.70 |
| COSTO OFICINA. 1 MES 4 personas  Wework, oficina por mes con todos los servicios  https://www.wework.com/es-LA/l/mexico-city--DIF |  | $19,400.00 |
| COMPRA CAÑON EPSON. SE RIFA AL FINAL DEL PROYECTO SI SE ENTREGA EN TIEMPO Y FORMA Fuente: Walmart |  | $9,299.00 |
| COSTO DEL PROYECTO |  | $98,703.70 |

Proyecto de BI, planeado, diseñad y ejecutado de manera estructurada

# 1. Kickoff del proyecto Inicio Planificación del proyecto.

El Kickoff fue una reunión en donde se presentó al equipo de trabajo con el cliente, se mostró el plan de trabajo, se platicó de que sería el proyecto y hubo café y galletas.

# Definir el proyecto.

## Robo de Identidad

“El Robo de Identidad, se define como la obtención ilegal de información personal de los usuarios de servicios financieros, con la finalidad de suplantarlos y, con ello, obtener recursos económicos, financieros o beneficios de forma no autorizada.

Las tendencias hasta 2017, señalan un crecimiento en el Robo de Identidad, de ahí que entidades nacionales están trabajando actualmente para legislar y definir los criterios de la autenticación biométrica para la banca nacional.” Fuente: presentación del Profesor sobre Robo de Identidad.

## Objetivo

Obtener un tablero del comportamiento del robo de identidad en México.

Revisar las tendencias de los últimos años del robo de identidad. Para fundamentar acciones que la mitiguen.

Verificar el planteamiento de las tendencias hasta 2017 del robo de identidad. Con un fundamento cuantitativo e independiente al reporte de Condusef.

## Alcance

Obtener una base de datos con base en las reclamaciones a la banca, información recabada por la Comisión Nacional Bancaria y de Valores (CNBV), e interpretada por la Condusef.

Realizar los procesos de transformación y limpieza para permitir el cargado de los datos.

Crear el modelo para obtener información que cubra los objetivos planteados.

## Limitaciones

No se cuenta con el reporte R27 generado por la CNBV. Se recabarán los datos con medios indirectos, que hacen uso de esta fuente.

Los datos utilizados serán los proporcionados por el gobierno federal.

Se utilizará PowerBI.

## Beneficios

Seguir los pasos de un proyecto de BI alcanzando los objetivos específicos del curso.

Proporcionar un modelo que permita el análisis del robo de identidad con los datos obtenidos.

# Definir la planificación y la gestión del proyecto.

## Plan de trabajo

El detalle de las actividades a realizar con estos roles, y el límite de tiempo se encuentra en el Excel adjunto, donde se desarrolla un diagrama de Gantt.



## Equipo de trabajo

El equipo de trabajo consiste en los siguientes roles:

|  |  |
| --- | --- |
| DP | Director de Proyecto |
| AC | Analista/consultor en Datawarehouse |
| CP | Consultor en Power BI |
| DB | Desarrollador en Business Objects y de ETL. |

El gobierno de TI para este proyecto es el siguiente:

## Director de proyecto.

**Perfil**.

Un profesional con conocimientos en marcos de trabajo y administración de proyectos. Estratégico. Con buen seguimiento a los procesos, capaz de comprender las tecnologías utilizadas en el desarrollo de servicios, capaz de coordinar esfuerzos y trabajo en equipo.

**Funciones**.

Generar las tácticas para los servicios de inteligencia de negocios que se proporcionan a la empresa, dirigidas a contar con conocimiento que apoye las estrategias de la empresa.

Proporcionar los recursos para ejecutar las tácticas y procesos operativos de los servicios de inteligencia de negocio de le empresa.

Realizar las juntas del equipo de BI.

Coordinar a su gente a cargo, o hacer seguimiento el proceso en caso de usar una metodología ágil.

Entrega ante cliente, interno o externo.

## Analista/consultor en Datawarehouse

**Perfil.**

Profesional en informática y área afines, con capacidad de análisis, conocimientos en diagramación lógica, algoritmos y estructuras de datos, desarrollo de software, lenguajes de programación. Con buena capacidad de auto capacitación y con gusto por las nuevas tecnologías. Trabajo en equipo y colaboración. Conocimientos Knowledge Discovery in Databases – KDD.

**Funciones.**

Actividades principales: Recibe y atiende las solicitudes de los servicios de Dtawarehouse de la prestación del servicio con miras a la mejora continua.

## Consultor en Power BI

**Perfil.** Profesional en informática y área afines, con capacidad de análisis, conocimientos en tableros de mando, KPI’s, dashboards y uso de PowerBI.

**Funciones.**

Diseña y construye el modelo en PowerBI que obtenga conocimiento adecuado, con respecto a los objetivos del proyecto y las estrategias de la empresa.

## Desarrollador en Business Objects y de ETL

**Perfil.**

Profesional en informática y área afines, con capacidad de análisis, conocimientos en diagramación lógica, algoritmos y estructuras de datos, desarrollo de software, lenguajes de programación. Con buena capacidad de auto capacitación y con gusto por las nuevas tecnologías. Trabajo en equipo y colaboración. Con conocimientos del proceso KDD y de ETL.

**Funciones.**

Recibe y atiende las solicitudes de los servicios de TI de procesos ETL. Construye software auxiliar en la extracción, transformación y limpieza de datos.

## Establecer la finalización del proyecto

La finalización del proyecto será al tener un modelo alterno al de Condusef, que tiene como fecha límite el 28 de mayo de 2018.

## 2.1. Revisar los requerimientos de negocio (usuarios, tiempos)

Al revisar los tiempos se concluye que el alcance será limitado a presentar el modelo en PowerBI y hacer una propuesta para el despliegue y mantenimiento.

## 2.2. Definir la arquitectura tecnológica (hardware)

Fuentes de datos externas

Imagen que contiene electrónica

Descripción generada con confianza alta

Internet



Las fuentes de datos serán externas y explotadas utilizando una conexión a Internet con una velocidad aproximada a la de la siguiente imagen:



Se usará una laptop Aspire F 15 con las siguientes características:

* Familia de procesador: Intel® Core™ i7-6xxx
* Diagonal de la pantalla: 15.6pulg.
* Memoria interna: 16 GB
* Tipo de memoria interna: DDR4-SDRAM
* Capacidad total de almacenaje: 1000 GB
* Sistema operativo instalado: Windows 10 Home7

## 2.3. Definir las recomendaciones de configuración

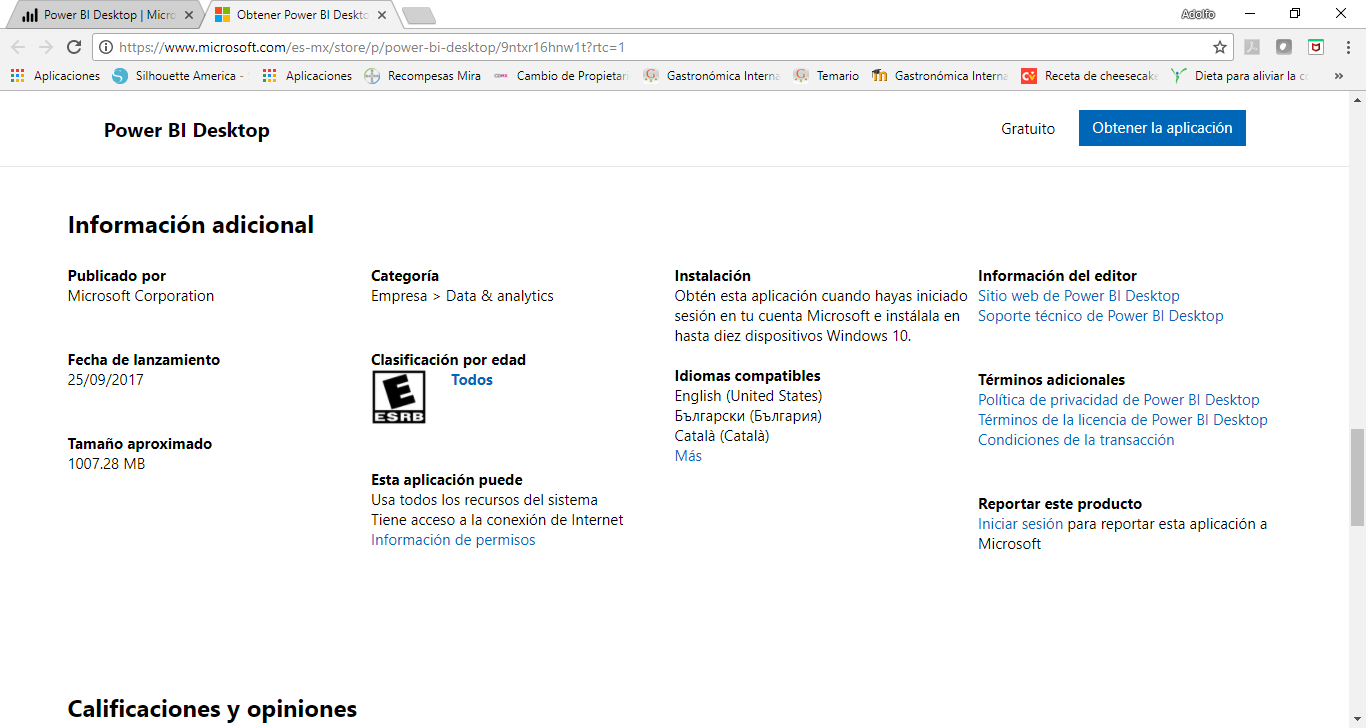
No se cuenta con recomendaciones espaciales de configuración, como cualquier otra solución funciona mejor si no compite con los recursos con *software*, por lo que se recomienda realizar las tareas relativas a este proyecto sin ejecutar programas que degraden el rendimiento de la computadora utilizada.

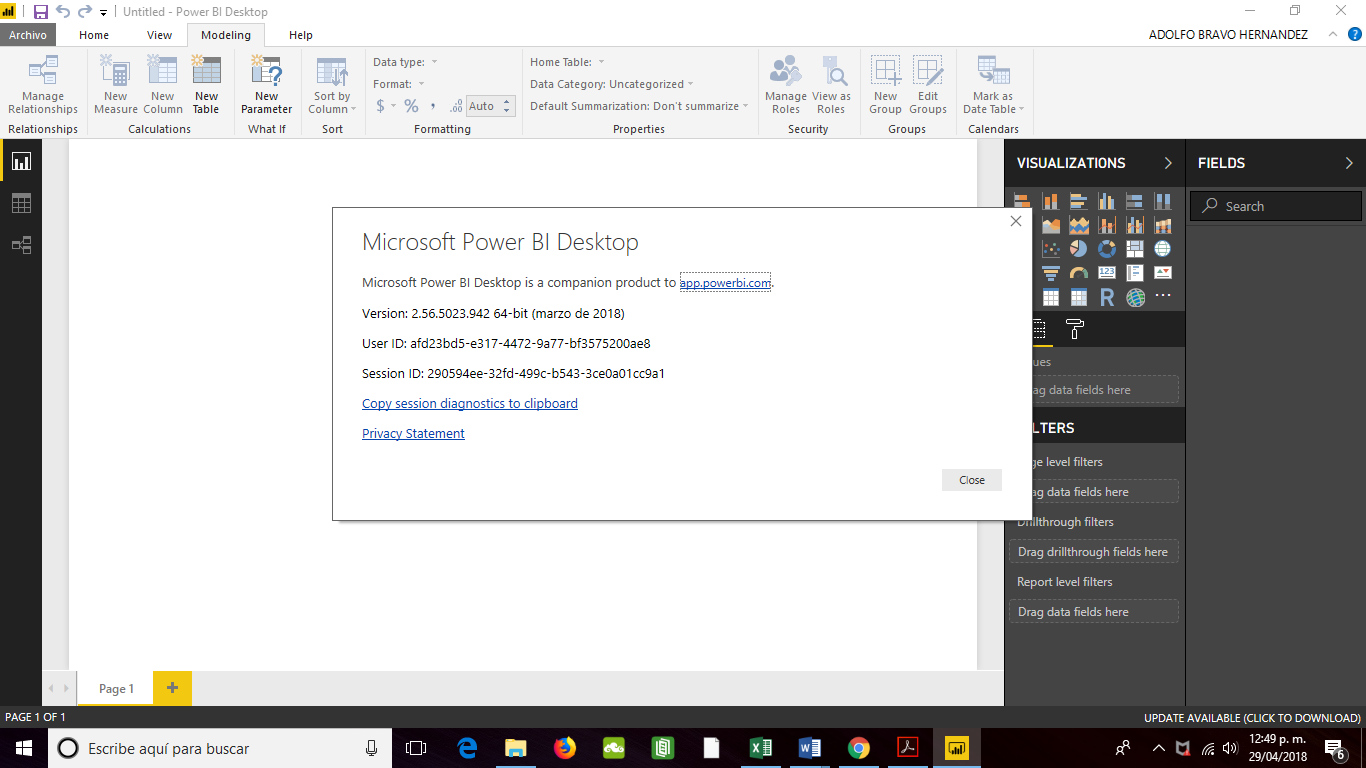
## 2.4. Estimar requerimientos de escalabilidad

La solución correrá bajo una computadora, presentando sólo un modelo en PowerBI, además los datos de insumo deben ir desplazando los más antiguos, por lo que no variaran significativamente en volumen, por lo que no se escalará.

## 2.5. Implementar el hardware y el software

Power Bi Desktop





Se instaló PowerBI en la computadora planeada, adicionalmente se usará Excel y Word (Office 365).

## 3.1. Analizar las fuentes de datos

Se considera como una premisa que existe una correlación bilateral entre la cantidad de reclamaciones que se hacen a los bancos por cargos no reconocidos y el robo de identidad. Con base en esto se obtendrán datos de dichas reclamaciones.

La Comisión Nacional Bancaria y de Valores (CNBV) genera un reporte 27 que contiene información de las reclamaciones hechas a los bancos, se genera de manera trimestral y tiene las siguientes consideraciones:

“Se entenderá por reclamación a todas aquellas operaciones monetarias no reconocidas por los clientes y que han sido comunicadas a la Institución por cualquier canal o medio puesto a disposición de los clientes de forma directa o indirecta.” (Tu gobierno en un solo punto., 2015)

### Fundamento legal del reporte

“Fundamento legal del reporte Las Entidades Financieras que en la prestación de Servicios de Inversión deberán enviar la información que se solicita en el siguiente reporte:

R27 B-2711 Reclamaciones Bursátiles

El envío de dicha información deberá efectuarse de acuerdo con lo señalado en las siguientes

disposiciones legales:

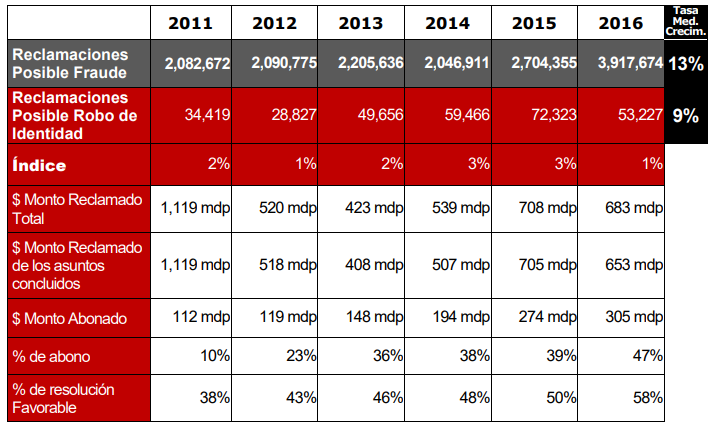
1. Facultades de la CNBV Ley de la CNBV, Artículo 4°, fracción V: “Corresponde a la Comisión: Expedir normas respecto a la información que deben proporcionarle periódicamente las entidades”.

2. De los Reportes Regulatorios y Medio de Envío Disposiciones de carácter general aplicables a las entidades financieras y demás personas que proporcionen servicios de inversión, publicadas en el Diario Oficial de la Federación el 9 de enero de 2015, modificadas mediante Resolución publicada en el propio Diario el 29 de julio, 26 de octubre de 2015 y 9 de febrero de 2016.

Artículo 51.- Las Entidades financieras (…) que proporcionen Servicios de Inversión deberán enviar a la Comisión la información que se menciona en el presente capítulo, mediante su transmisión vía electrónica, utilizando el SITI.” (Tu gobierno en un solo punto, 2016)

### Datos del reporte

El gobierno no expone directamente el reporte 27, este es analizado por Condusef y muestra información al respecto como la siguiente (Condusef, 2017) :



Lo cual es una tabla con información bastante difícil de superar, pero estamos atenidos a el análisis de dicha institución, la idea en el presente proyecto es encontrar conocimiento, pero basado en datos del reporte original. El cual al no ser expuesto directamente se extraerá de manera indirecta haciendo consultas a través de un Excel que proporciona el mismo gobierno y que tiene una conexión de datos a una fuente externa.

El Excel que expone el gobierno tiene los siguientes datos (Portafolio de información CNBV, 2016):

|  |  |
| --- | --- |
| Usuario | Anónimo |
| Sector financiero | banca Múltiple |
| Tema | INFORMACIÓN OPERATIVA |
| Reporte | 040\_27A\_R3 |
| Nombre del reporte | R3. Reclamaciones por producto financiero, canal, motivo de reclamación, causa de resolución y estatus de la reclamación |
| Liga | http://portafolioinfdoctos.cnbv.gob.mx/Documentacion/minfo/040\_27a\_R3.xls |

Dicho archivo se encuentra en el siguiente adjunto:





## 3.2. Desarrollar los modelos de datos

Con base en los datos que se muestran en el archivo se generó el siguiente modelo de datos:

Imagen que contiene captura de pantalla

Descripción generada con confianza muy alta

## 3.3. Diseñar la base de datos

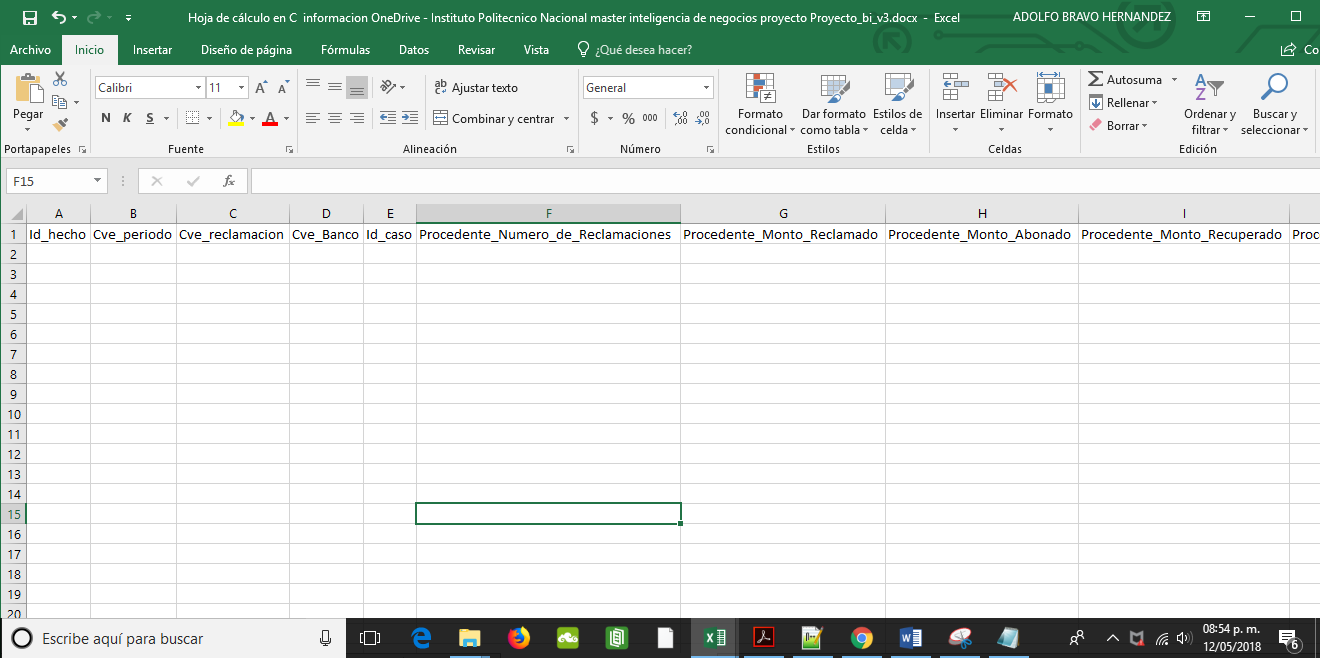
La base de datos será el conglomerado de las consultas usando el archivo 040\_27a\_R3.xls. y creando una hoja por cada una de las tablas, las cuales serán exportadas a PowerBI. Esto bajo la consideración que el reporte es trimestral y que actualizar los datos se hará de manera poco frecuente, probablemente una vez cada tres meses.

## 4.1. Revisar el alcance y la planificación

El alcance será encontrar conocimiento con base en los datos proporcionados por el gobierno. La limitante será que se tiene un medio indirecto para obtener los datos, lo que podría generar diferencias en lo encontrado por Condusef.

## 4.2. Implementar la base de datos

Se creó el Excel como se propuso. Junto con una hoja para el **diccionario de datos**.



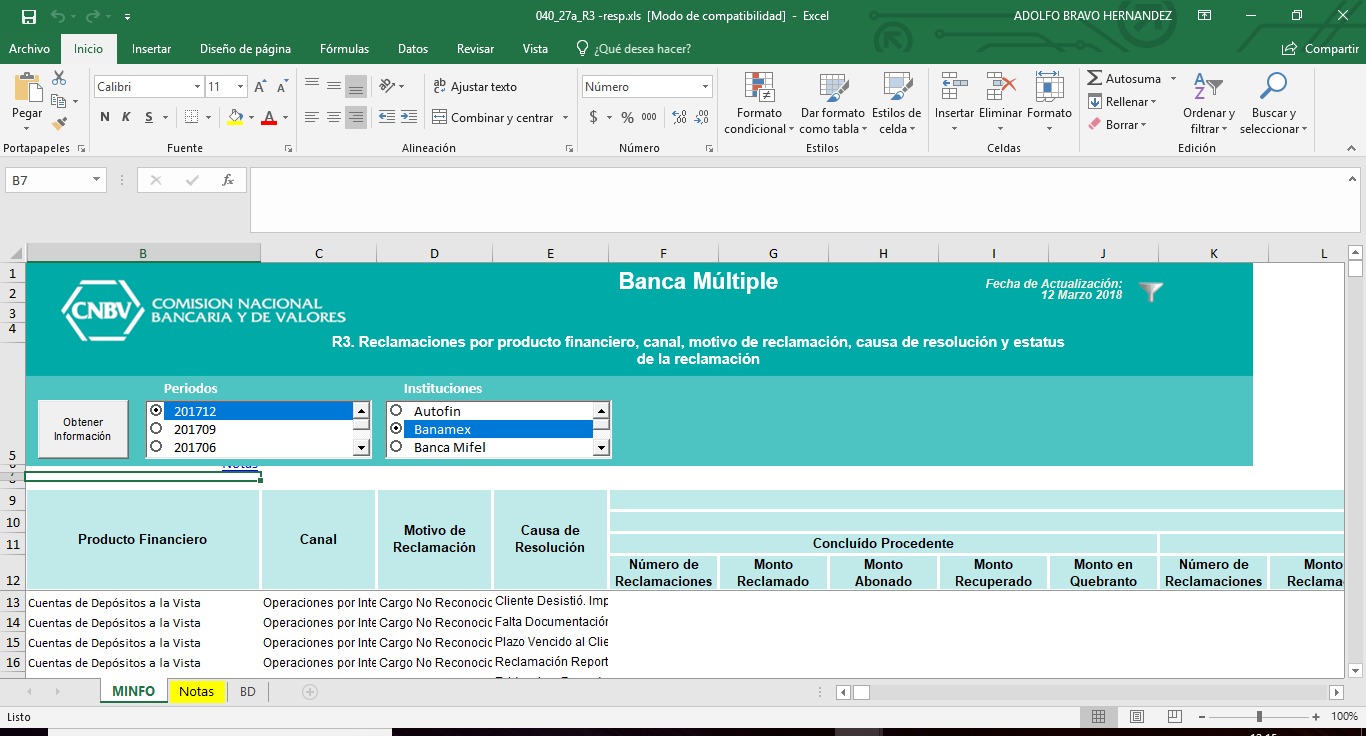
Adjunto a continuación.

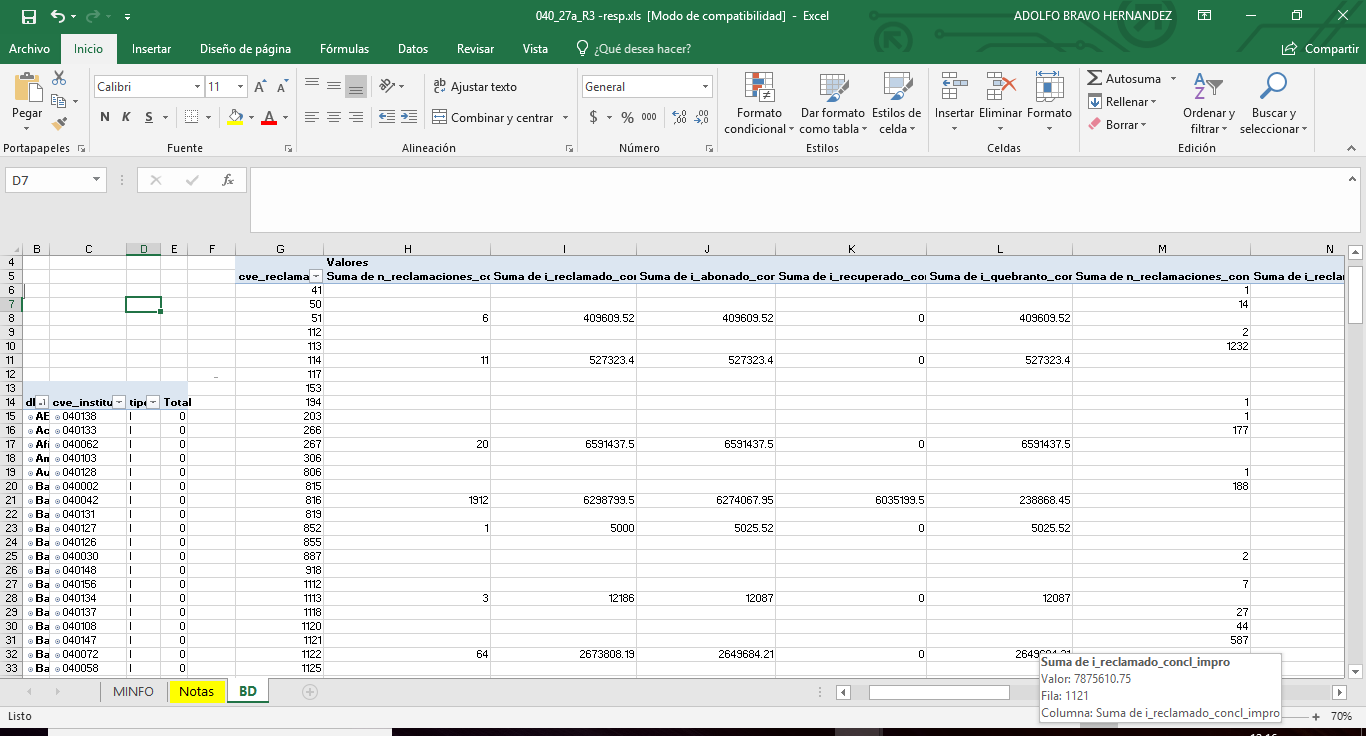


## 4.3. Diseñar y desarrollar la integración de datos

Se aglomeran los datos obtenidos desde el Excel 040\_27A\_R3.

Ejemplo de consulta:



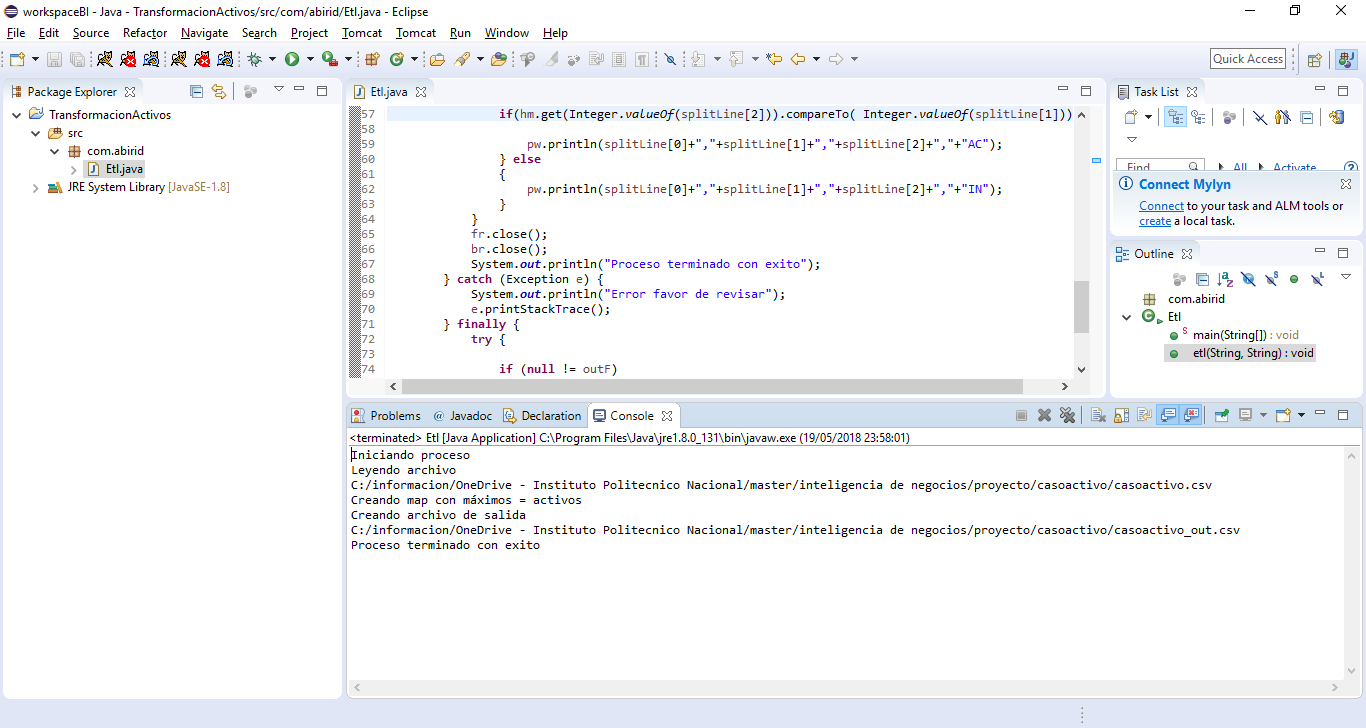


## 4.4. Cargar y validar la base de datos (en progreso)

Se realiza el siguiente proceso de transformación y limpieza:

1. Todos los datos blancos se cambian por cero por ser congruente con el tipo de dato, por facilitar los cálculos y no generar “basura” para los datos.
2. La dimensión tiempo se obtiene extrayendo un campo del origen de datos la fecha de la reclamación que viene en el formato YYYYMM de cada reclamación, es decir año y mes.
   1. Se quitarán los duplicados
   2. Se separará el dato en YYYY y MM, con las fórmulas =EXTRAE(celda,1,4) y =EXTRAE(,1,4), después de lo cual se hace un pegado especial de valores para quitar la formula y dejar sólo el valor.
3. Reclamación, para obtener esta hoja simplemente se quitaron los renglones duplicados con las herramientas de Excel de “Quitar duplicados”.
4. Para generar la hoja Estado\_caso, se colocará como AC el caso más reciente de un número de reclamación.

Cuando una reclamación es abierta puede permanecer un tiempo en los reportes, por lo que la clave de la reclamación aparece varias veces al obtener los datos. Para obtener cual es el renglón más reciente de cada número de reclamación, que además puede aparecer en más de un banco en un mismo periodo, se construirá un programa en java que haga esta extracción.



Se adjunta archivo jar con el código fuentes incluidos.



## 4.5. Crear modelo en Power BI

## 5.1. Entregar la documentación del proyecto

Milestone a concluir el 28 de mayo de 2018. Consistirá en entregar de manera impresa el reporte y el plan de proyecto, con evidencias del despliegue del modelo.

## 5.3. Entregar modelo y conclusiones

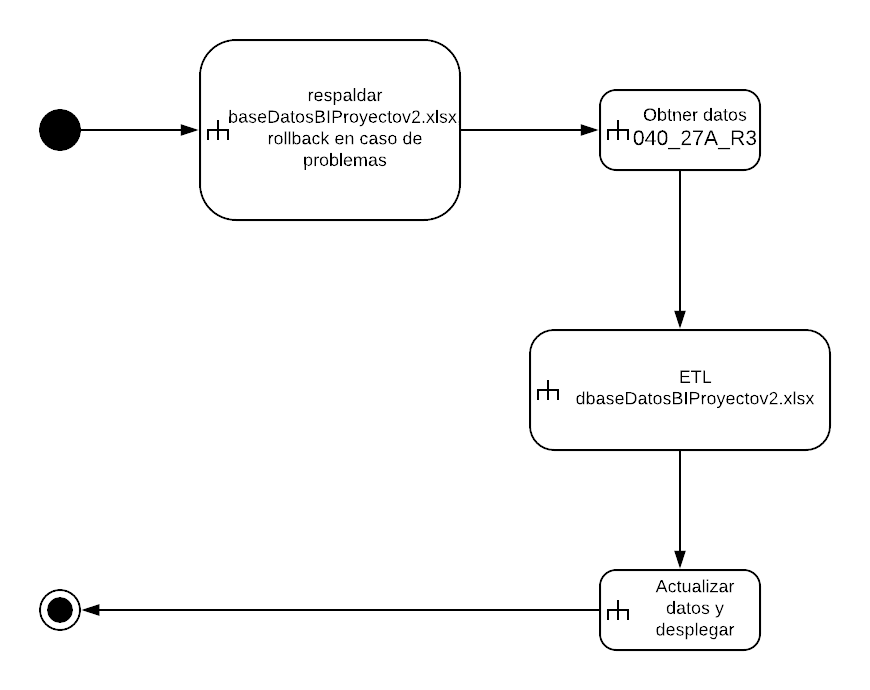
## 6.1. Propuesta para operar, mantener y mejorar la aplicación

Operación:

Debido a que los datos no cambian continuamente, la operación es sólo revisar periódicamente que el modelo se encuentre disponible.

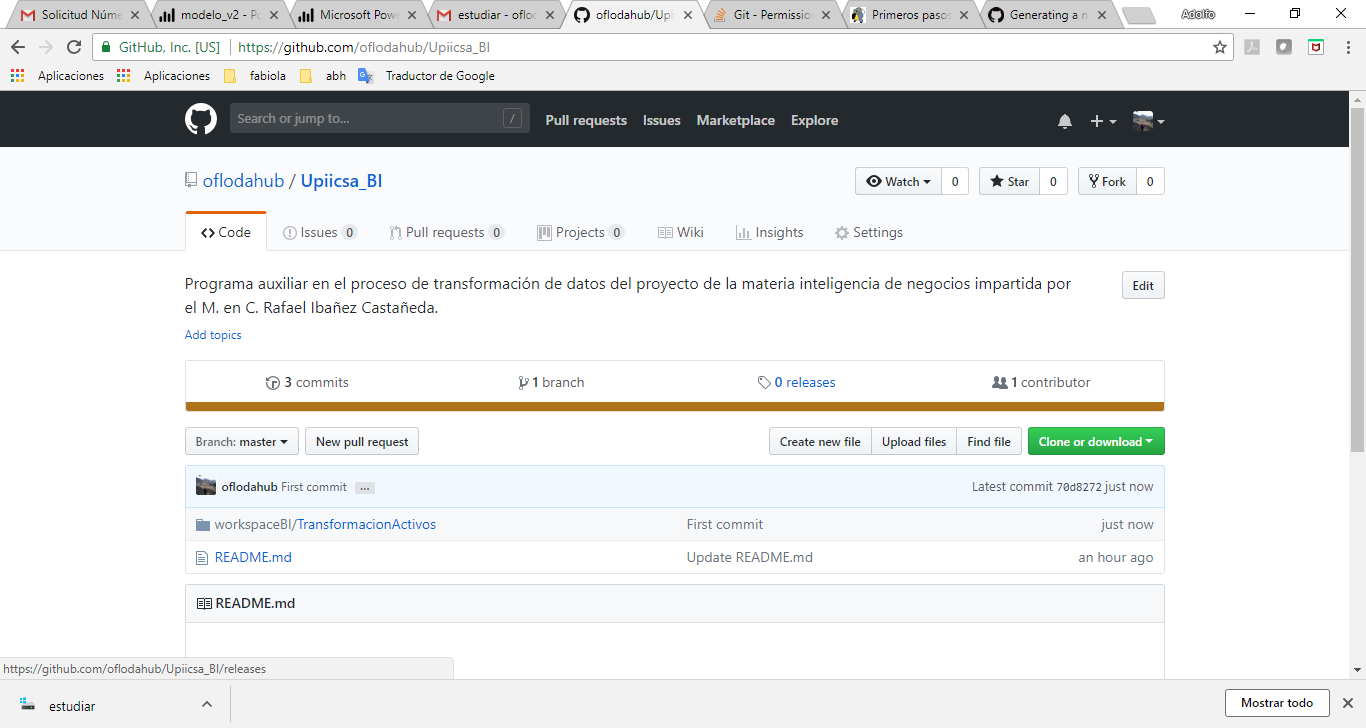
### Mantenimiento:

* Se requieren hacer consultas por banco y nuevo periodo con el Excel 040\_27A\_R3
* Los datos se modifican de manera que los blancos se ponen en cero
* Se respalda el Excel que se da a continuación:



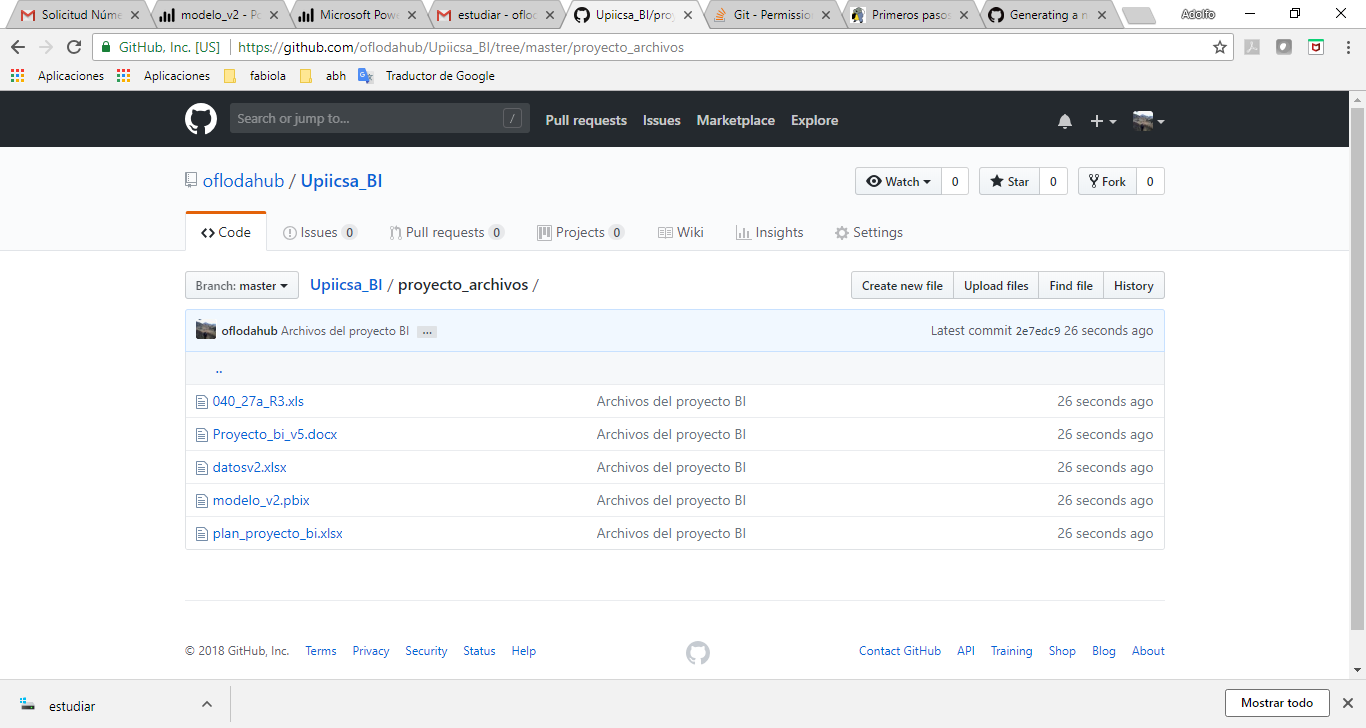
* Se colocan los datos en layout como sigue
  + Hoja Reclamación se agregan de las columnas: Cve\_reclamacion Producto\_Financiero Canal Motivo\_de\_Reclamacion Causa\_de\_Resolucion
  + Tabla\_de\_hechos\_reclamaciones se agregan las columnas: Id\_hecho Cve\_periodo Cve\_reclamacion Cve\_Banco Id\_caso Procedente\_Numero\_de\_Reclamaciones Procedente\_Monto\_Reclamado Procedente\_Monto\_Abonado Procedente\_Monto\_Recuperado Procedente\_Monto\_en\_Quebranto Improcedente\_Numero\_de\_Reclamaciones Improcedente\_Monto\_Reclamado Improcedente\_Monto\_Abonado Improcedente\_Monto\_Recuperado Improcedente\_Monto\_en\_Quebranto Pendiente\_Numero\_de\_Reclamaciones Pendiente\_Monto\_Reclamado Pendiente\_Monto\_Abonado Pendiente\_Monto\_Recuperado Pendiente\_Monto\_en\_Quebranto
  + Para la Hoja Estado Caso, se tiene que ejecutar el programa en java que se encuentra en:

<https://github.com/oflodahub/Upiicsa_BI>



* + De la hoja Tabla\_de\_hechos\_reclamaciones se copian las columnas
  + Id\_hecho Cve\_periodo Cve\_reclamacion Cve\_BancoDel a un CSV, se ejecuta el programa y se copian las columnas de la salida 1 y 4 en la hoja de Estado.
* Se actualiza el modelo en el archivo modelo\_v2.pbix que se encuentra en:

<https://github.com/oflodahub/Upiicsa_BI/tree/master/proyecto_archivos>



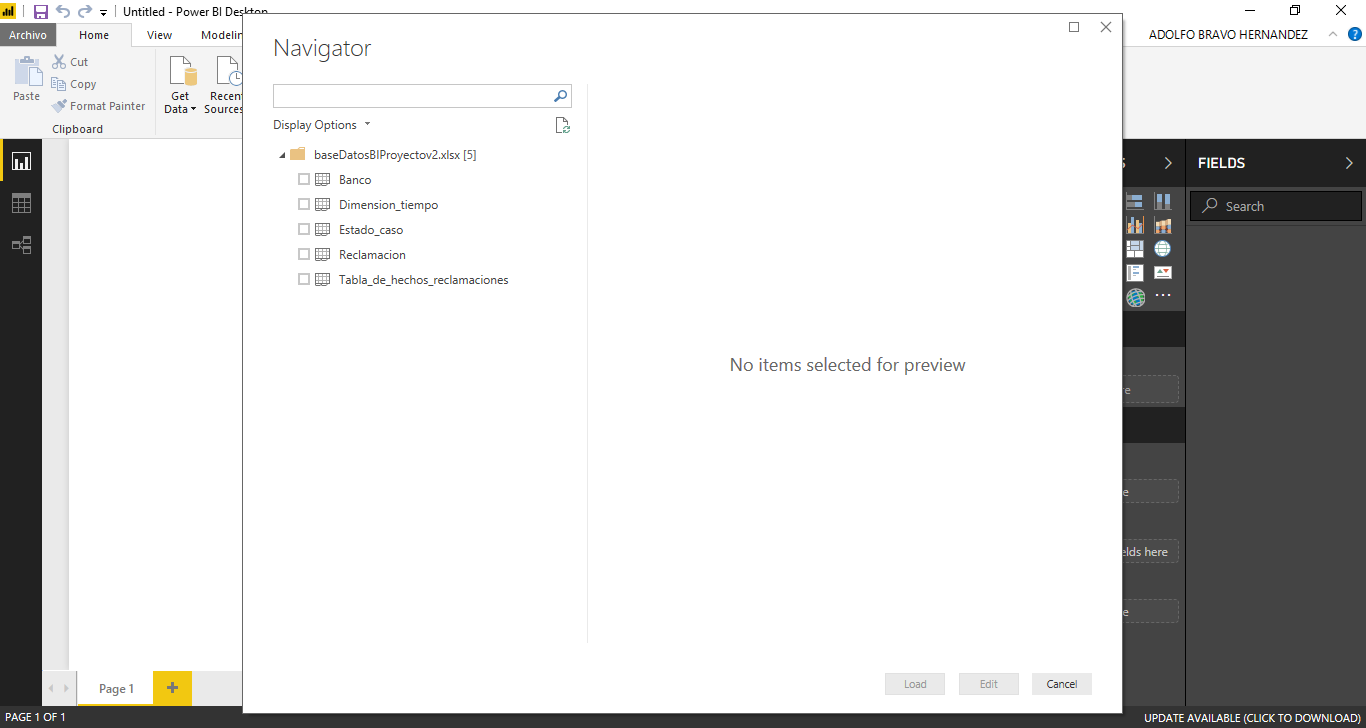
### Mejoras:

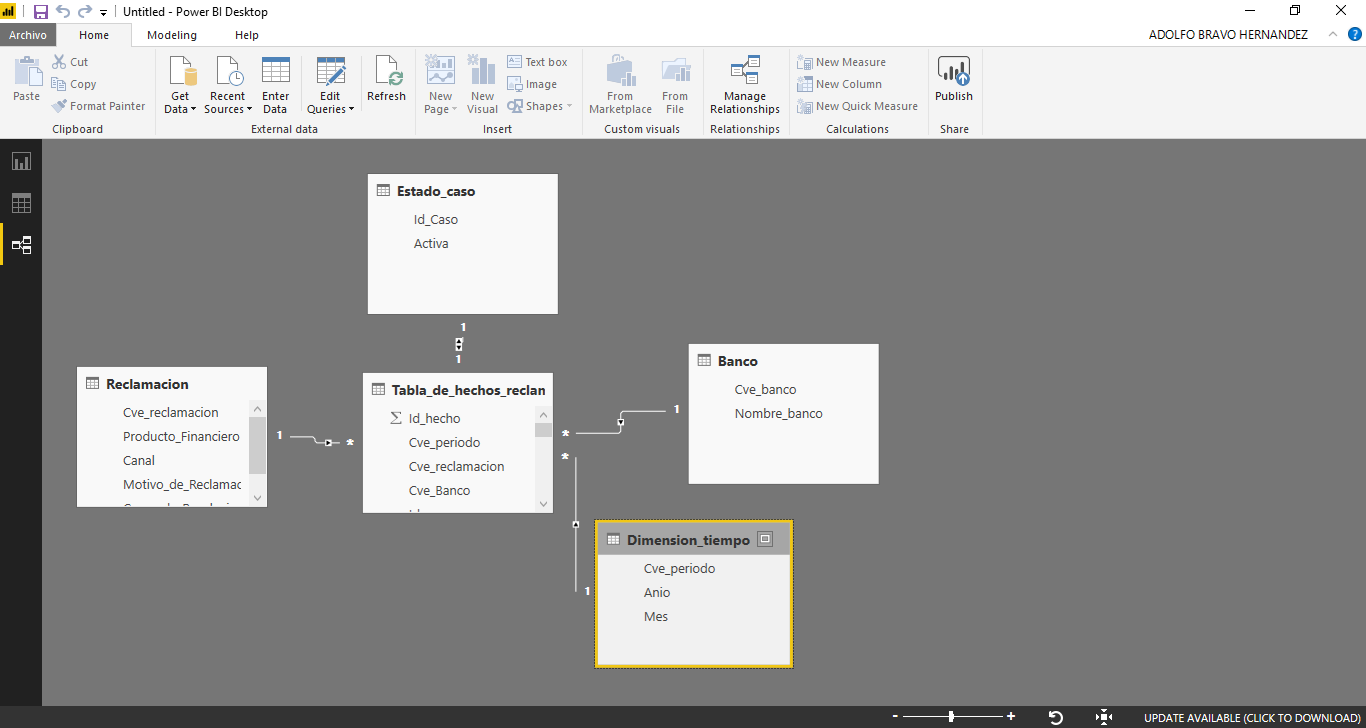
Como la mayor parte de los modelos, la mejora estará en mejorar la cantidad de variables y adquirir datos más completos del reporte 27, es un poco complicado pues estos datos suelen no ser expuestos de manera pública de manera directa, o no se ha difundido dicho acceso.

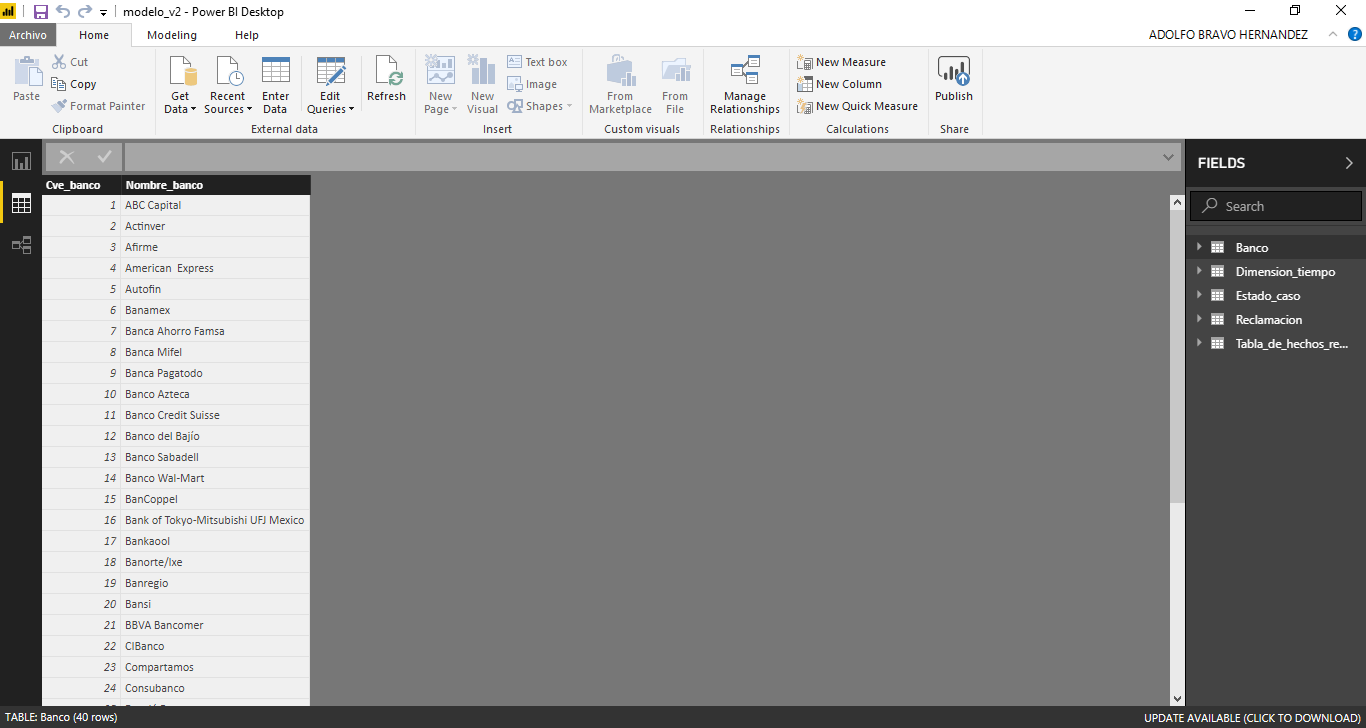
# Modelo

Se recuperan los datos que se tienen en el archivo baseDatosBIProyectov2.xlsx que se encuentra en el repositorio GIT en la ruta:

<https://github.com/oflodahub/Upiicsa_BI/tree/master/proyecto_archivos>



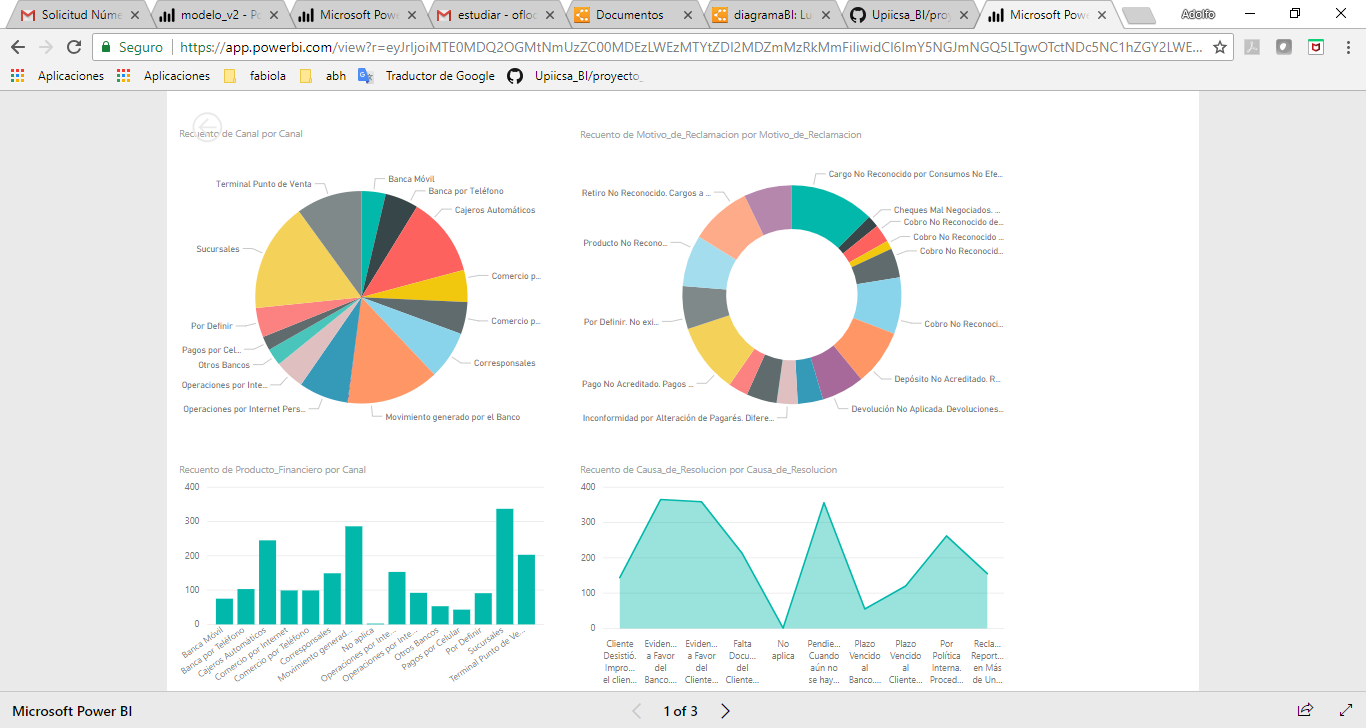


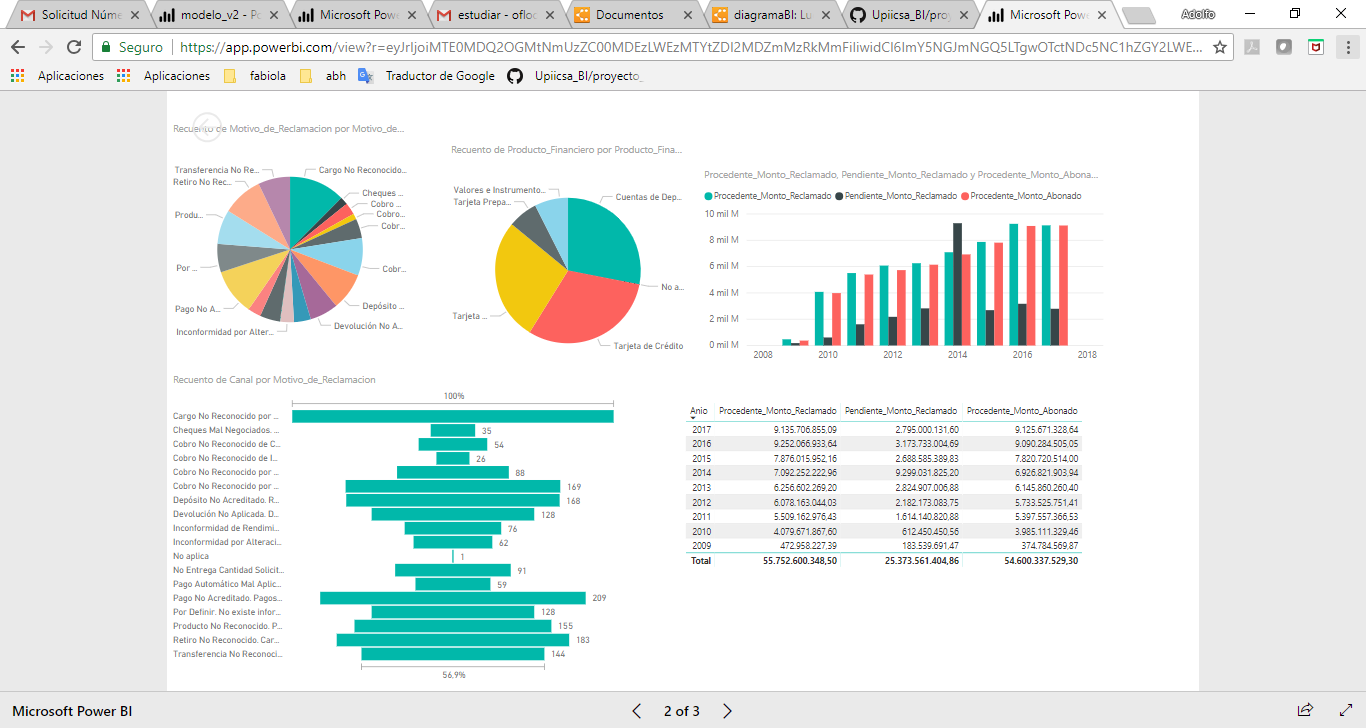


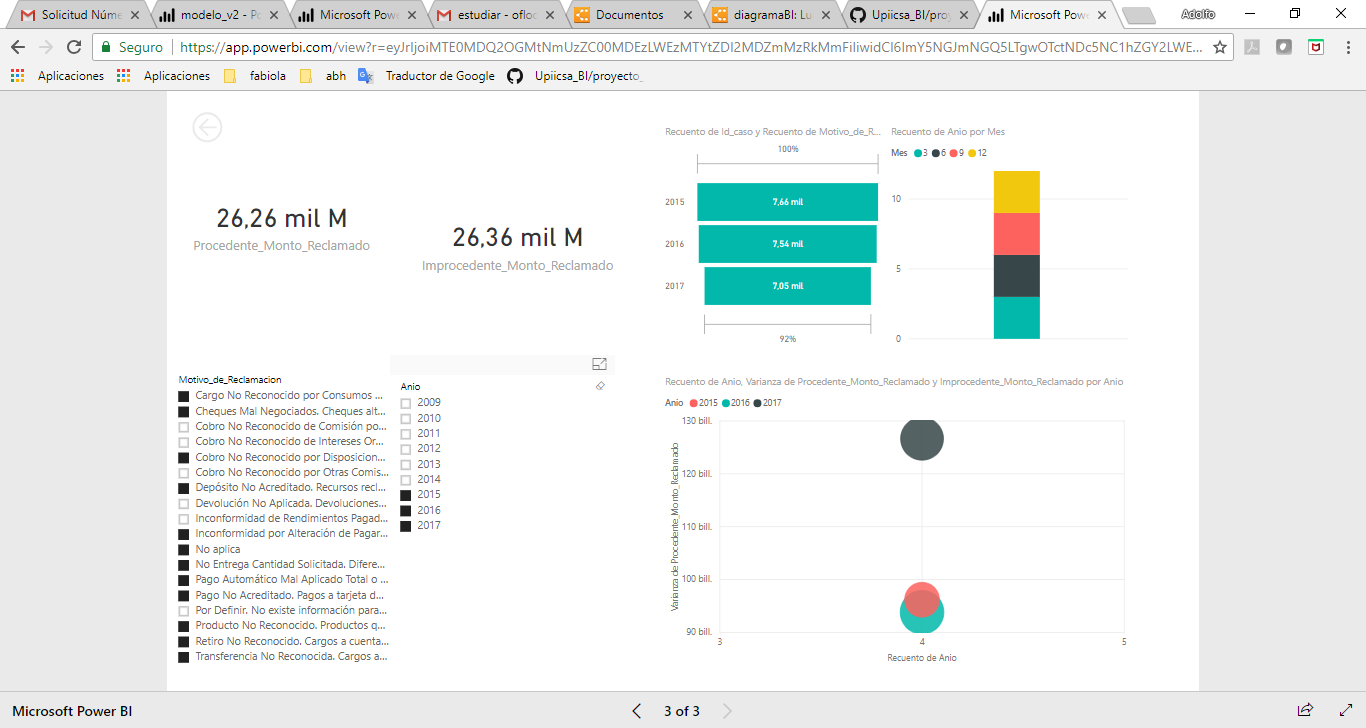
* A partir de ahí se creó el modelo que se encuentra modelo\_v2.pbix que se encuentra en:

<https://github.com/oflodahub/Upiicsa_BI/tree/master/proyecto_archivos>

El modelo es desplegado en la ruta:





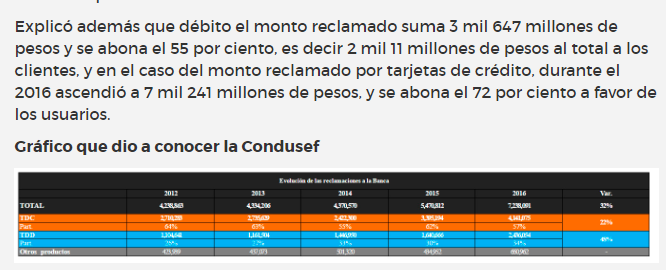


# Conclusiones:

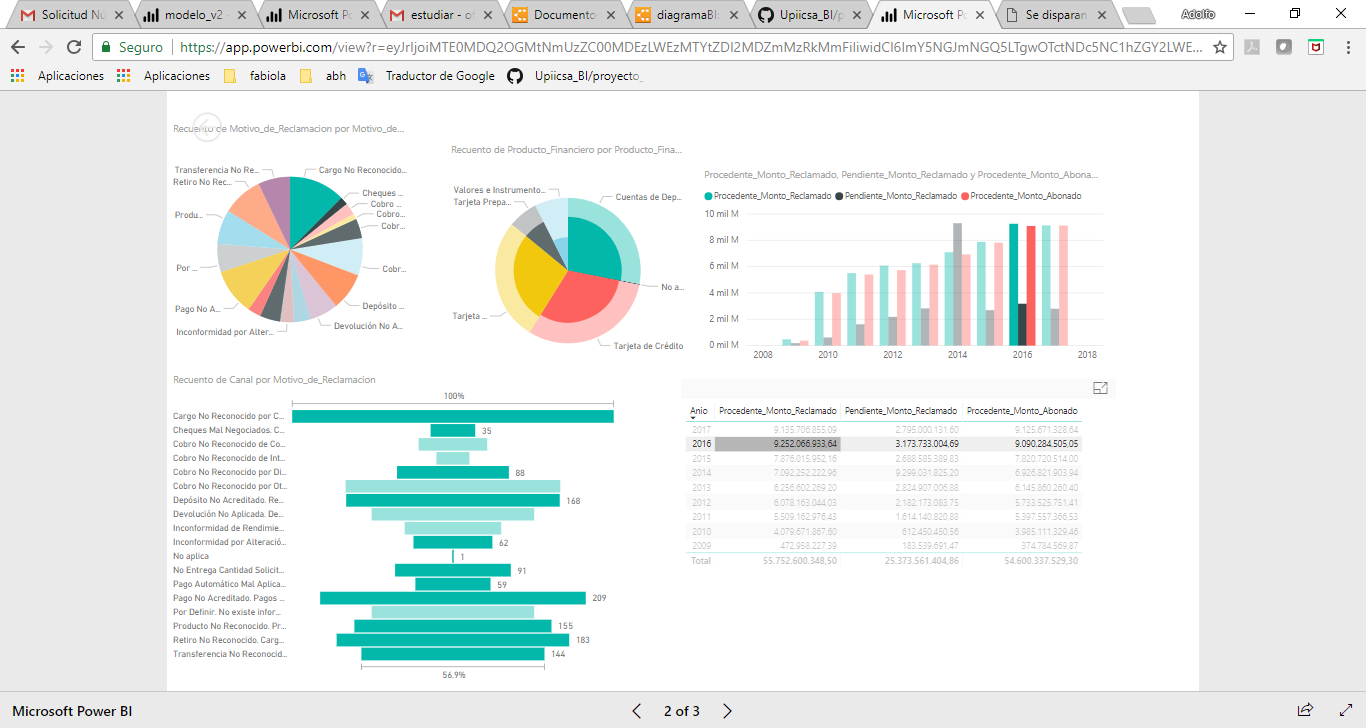
* Hay reportes que nos dan una idea como el de la página:

<https://mexico.quadratin.com.mx/se-disparan-los-reclamos-a-los-bancos-por-fraudes-con-tarjetas-de-debito/>

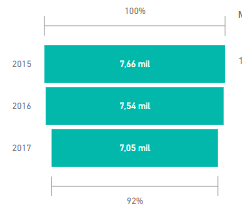
Que dice:

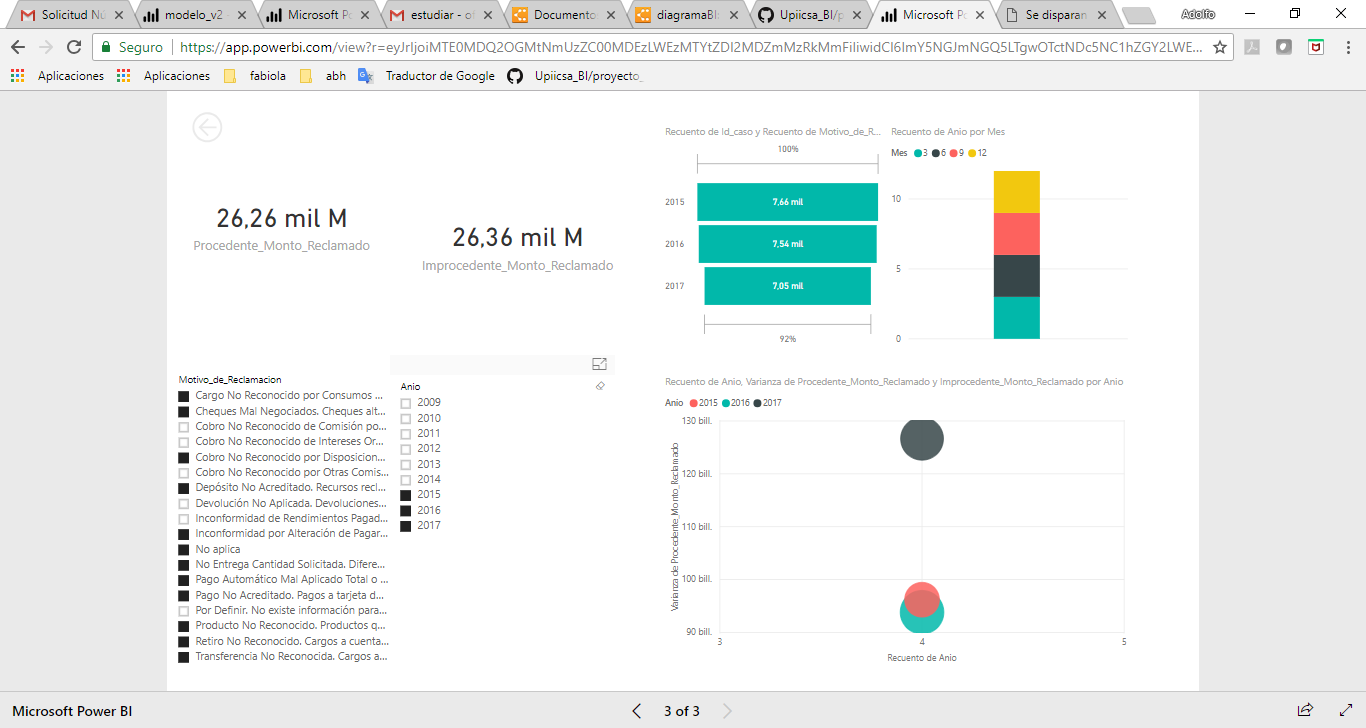


Pero son un tanto cuanto parciales, por ejemplo cuando con los datos proporcionados efectivamente el monto total de 2016 es de cerca de 9 mil 256 millones de pesos en total, pero con conceptos sumados con una clasificación pedestre como el de “No aplica”.

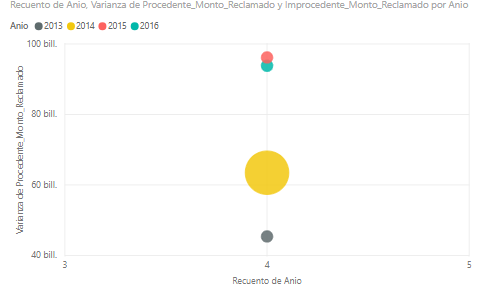


* Con los datos proporcionados hay una baja en las reclamaciones, por lo que a nivel de indicador significa que los controles que se hayan implementado para controlar el Robo de Identidad tuvieron un efecto positivo para su control.

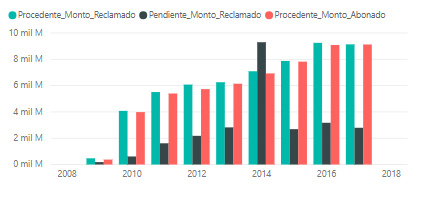




* En 2014 fue un año con una gran variación en los montos reclamados. Se ve una bola amarilla mucho más grande comparada con otros años debido a la enorme varianza.



* En 2014 además hubo una gran cantidad de Montos Pendientes, seguramente no ha de ver sido nada agradable estar en procesos de reclamación, pues muy seguramente ha de ver sido más lento de lo común.



* Los procesos de inteligencia de negocio son aplicables en prácticamente cualquier proceso donde podamos recabar datos. Es conveniente que un experto en el tema (del dominio) participe para poder ajustar el modelo.
* Se puede obtener conocimiento a partir de datos de manera estructurada. Es mucho más claro y ordenado si éste proceso se realiza utilizando una metodología.
* Los modelos son elementos que deben constantemente afinarse, y cambiarán con la una frecuencia semejante a la evolución de los datos que lo originen.
* Contar con datos de poca calidad complica el proceso de creación de modelo. Se pueden realizar procesos de limpieza y transformación, pero estos tienen un límite.
* Varios de los procesos gubernamentales no cuentan con la calidad de datos e información que deberían si se invirtiera más en el proceso. Incluso pareciera que no tienen un proceso plenamente consolidado.
* Repasando los objetivos:
  + Obtener un tablero del comportamiento del robo de identidad en México.

**Se cumple con el modelo desplegado.**

* + Revisar las tendencias de los últimos años del robo de identidad. Para fundamentar acciones que la mitiguen.

**El robo de identidad reflejado por la relación con las reclamaciones a la banca, han ido a la alza, sin embargo en 2017, hubo una pequeña baja, por lo menos en este rubro las medidas que se utilizaron fueron efectivas.**

* + Verificar el planteamiento de las tendencias hasta 2017 del robo de identidad. Con un fundamento cuantitativo e independiente al reporte de Condusef.

**Yo de manera personal esperaba encontrar una alza de la cantidad de reclamaciones en 2017. El objetivo se cumplió al encontrar la tendencia con base en los datos proporcionados.**

# Referencias

Condusef. (26 de 01 de 2017). *Tu gobierno en un solo punto*. Recuperado el 12 de 05 de 2018, de gob.mx: https://www.gob.mx/cms/uploads/attachment/file/185241/r27-3erTrim2016ReclamImpactoMonetario.pdf

*Portafolio de información CNBV*. (19 de 11 de 2016). Recuperado el 12 de 05 de 2018, de INFORMACIÓN OPERATIVA: http://portafoliodeinformacion.cnbv.gob.mx/Lists/ConsultasRegistradas/DispForm.aspx?ID=544861&ContentTypeId=0x0100A7129E0DCEB0FE4CA0D8E6AD4D769E8E

*Tu gobierno en un solo punto*. (12 de 12 de 2016). Recuperado el 12 de 05 de 2018, de gob.mx: https://www.gob.mx/cms/uploads/attachment/file/174446/DCGEF\_y\_AI\_\_R27\_B-2711\_Reclamaciones\_Burs\_tiles\_201612.pdf

*Tu gobierno en un solo punto.* (30 de 09 de 2015). Recuperado el 12 de 05 de 2018, de gob.mx: https://www.gob.mx/cms/uploads/attachment/file/121850/BM\_BD\_R27\_A-2701\_Reclamaciones\_Periodicidad\_201601-actual\_Actualizaci\_n\_01\_de\_Septiembre\_de\_2015.pdf