



Máster en Business Intelligence y Big Data on Line EOI – 2018/2019

PROYECTO DE MEJORA EN LA GESTIÓN DE COBRANZAS

HITO 6. MEMORIA

Mejora en la gestión de cobranzas Hito

6. MEMORIA

Contenido

1	Introducción	5
1.1	Antecedentes	5
1.2	Quiénes somos	7
2	Hipótesis.....	8
2.1	Correlación entre envíos y cobranzas	8
2.2	Identificación de los motivos que generan impagos.....	8
2.3	Optimización de rutas	9
3	Oportunidad de negocio	9
4	Supuestos de usuario	10
5	Fuentes de datos	11
5.1	Encuestas.....	11
5.1.1	Cobranzas	11
5.1.2	Finanzas	14
5.1.3	Business Intelligence	16
5.2	Base de datos corporativa	17
5.3	Datos generados y de terceros.....	21
6	Hipótesis de trabajo	22
7	Análisis y diagnóstico	25
7.1	Situación actual	25
7.2	Debilidades, amenazas, fortalezas y oportunidades.....	26
8	Modelo de Negocio	28
8.1	Key Partners	31
8.2	Key Activities	31
8.3	Key Resources	32
9	Plan de acción	32

9.1	Alcance, objetivos y métricas.....	33
9.1.1	Ratio de visitas duplicadas por cliente	34
9.1.2	Reducción del tiempo dedicado a la planificación de rutas.....	37
9.2	Actividades: Mapa de procesos y solución tecnológica.....	38
9.2.1	Mapa de procesos	38
9.2.2	Solución tecnológica.....	39
9.2.3	Arquitectura técnica.....	40
9.3	Organización y recursos	48
9.4	Gestión del tiempo.....	48
9.4.1	Definición del modelo de negocio y planificación.....	48
9.4.2	Desarrollo	49
9.4.3	Despliegue.....	49
9.4.4	Entrenamiento continuo y aprendizaje.....	49
9.4.5	Desarrollo de negocio	49
10	Proyecto de creación.....	51
10.1	Plan de inversión	51
10.1.1	CÁLCULO DE COSTES	52
10.1.2	POC.....	52
10.1.3	Fase de implementación	52
10.1.4	Proyecto Fase de soporte.....	53
11	Proyecto de optimización.....	53
11.1	Beneficios tangibles.....	53
11.1.1	Reducción de costes	53
11.2	Beneficios intangibles.....	55
11.3	Análisis económico-financiero para el Cliente	55
12	Flujo de Caja OPTIMIZA	57
13	Programa detallado de recursos (en precio de venta).....	58
14	Conclusiones.....	63
	Índice de ilustraciones.....	64
	ANEXO	65
	Anexo I: Código de generación del dataset del modelo.	65

Anexo II: Código del modelo supervisado.....	68
Modelo de predicción de cobranzas para Venado-Testing¶.....	68
ETL y Análisis descriptivo¶.....	68
Planificación de rutas¶.....	78

1 Introducción

OPTIMIZA nace como un servicio de consultoría orientado a la modernización de procesos, que trabaja para mejorar la gestión de las cobranzas del Grupo Venado en Bolivia.



Ilustración 1. Localización. Fuentes: Wikipedia, Grupo Venado

1.1 Antecedentes

Grupo Venado se dedica principalmente a la fabricación y comercialización de productos alimenticios en general. Asimismo, también tienen como actividad el procesamiento y comercialización de productos destinados al cuidado del hogar y cuidado personal.

- El Grupo nace de la fusión, realizada en el 2015, entre Industrias Venado S.A, Fábrica Cruceña de Levaduras SA, Industria de Alimentos Santander S.A y Distribuidora Santa Cruz S.A. Cabe resaltar que Industrias Venado S.A fue fundada en 1912 en La Paz, Bolivia.
- Actualmente tiene 4 fábricas en diferentes lugares del país: una en La Paz, una El Alto, una en Warnes y una en Santa Cruz.
- Cuenta con más de 150 productos de las marcas Kris, Bristar, Fleischmann, Real, Frussion, Tolon.
 - Kris ofrece salsas (mayonesa, mostaza, ketchup y salsa golf), puré de papas, aceite de oliva, refrescos en polvo, entre otros.
 - Bristar ofrece productos de limpieza.
 - Real ofrece salsas (básicamente aderezos).
 - Fleischman ofrece levaduras y polvos para hornear.
 - Frussion ofrece jugos
 - Tolon ofrece yogurt

El portafolio de productos de la Empresa contempla 25 líneas de negocio. De ellas, 4 concentran el 83% de las ventas: Salsas (65%), Levaduras (9%), Postres en polvos (5%), Polvos para hornear (4%), facturando a la fecha 80 millones de USD\$ anuales o 559 millones de BOB anuales.

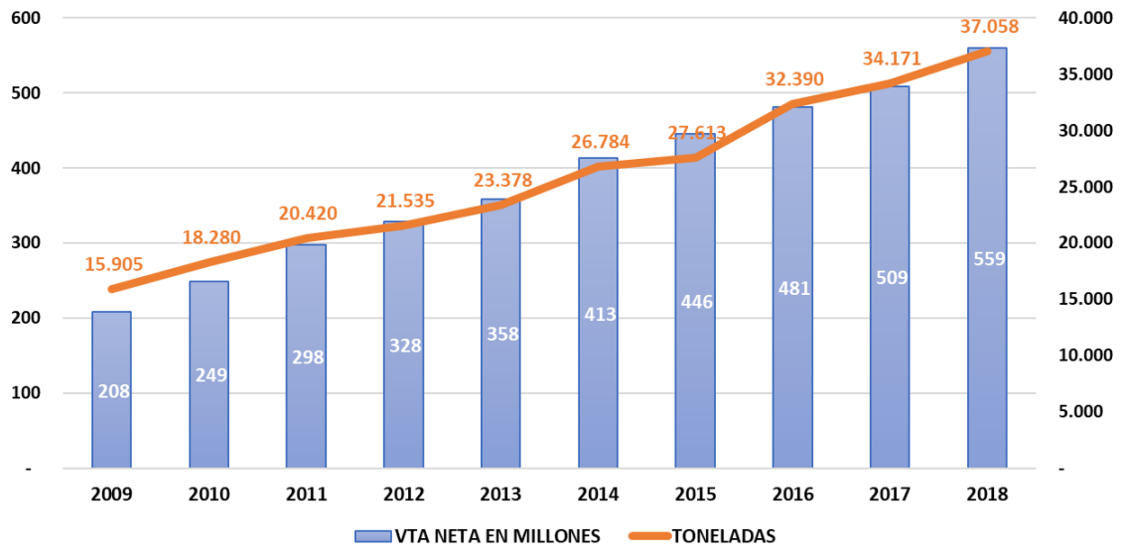


Ilustración 2. Volumen actividad Venado. Fuente Informe Anual Grupo Venado

Bolivia no es un país muy bancarizado. Adicionalmente la implantación de medios de pago digital en el país no supera el 15%.

Si analizamos el tipo de clientes de la empresa, al ser productor de alimentos que apuntan a:

- los vendedores de los mercados,
- las tiendas de barrio,
- supermercados,
- restaurantes

El 86% de la venta realizada es cobrada en efectivo por lo que, para realizar los cobros, la empresa dispone de dos grupos de cobradores: los que van a pie y los choferes-cobradores. Se cuenta con esta fuerza en 8 departamentos de 9 a nivel nacional, todos excepto Pando.



Ilustración 3. Bolivia. Fuente: Maquetas Santa Cruz

El proyecto consiste en desarrollar una herramienta para Grupo Venado que le permita mejorar este proceso con los recursos de los que dispone y de esta manera disponer de mayor liquidez para el desarrollo del negocio.

1.2 Quiénes somos



David – Machine Learning



Susana – Financiero



Sandra – Desarrollo



Marco – Base de Datos



Alfredo – Operaciones

2 Hipótesis

2.1 Correlación entre envíos y cobranzas

Creemos que el cliente tiene un problema con la correlación entre las ventas, la entrega de mercancía y los cobros.

Actualmente el departamento de cobros es el encargado de identificar qué clientes son susceptibles de ser cobrados pero la realidad dice que en un alto porcentaje de casos el cliente todavía no ha recibido la mercancía. La principal razón es la ausencia del cliente cuando se produce el envío programado. Esto produce un alto ratio de impagos por mercancía no entregada, aunque esto solo ocurre cuando los cobros los realiza la fuerza de cobradores a pie. En el caso de los choferes cobradores esto realmente no pasa, ya que los encargados de entregar la mercancía también se ocupan del cobro.

En estos momentos no existe una herramienta que informe tiempo real de la ejecución del envío.

Por otro lado, el departamento de cobros realiza manualmente las comprobaciones de ventas y entregas para elaborar sus rutas de cobro. Este proceso tiene un coste humano elevado y es susceptible a errores, impactando directamente en la liquidez del flujo de caja, gastos de desplazamiento de cobradores y malestar del cliente.

Además, hay 3 personas que dedican más de una hora de su trabajo a correlar las ventas y planificar las rutas de cobros de manera diaria y manual. Un proceso que es poco eficiente y muy dependiente de esas personas en concreto, que son las únicas que tienen el conocimiento y la experiencia.

2.2 Identificación de los motivos que generan impagos

Creemos que el cliente tiene un problema para clasificar los impagos y conocer el motivo por el que se producen.

Se han identificado las siguientes causas por las que los clientes del Grupo Venado suelen incurrir en impagos:

- Mercancía no servida.
- Establecimiento del cliente cerrado.
- Horario de visita del cobrador inadecuado.
- Falta de voluntad de pago / no conformidad del cliente.
- Fraude entre cliente, vendedor y cobrador.

El equipo de cobranzas elabora manualmente las rutas de cobro sin tener en cuenta estas variables.

Haciendo un estudio específico de estos valores se creará un perfil de comportamiento de cada cliente, lo que posibilitará un incremento de aciertos en los cobros por visita.

2.3 Optimización de rutas

El cliente elabora rutas de forma manual y tradicional sin tener en cuenta factores internos y externos a Grupo Venado.

Actualmente, dentro de Grupo Venado existe el equipo de supervisores de cobros que diariamente analizan las rutas pasadas y elaboran periódicamente las futuras. Para ello se tienen en cuenta dos factores principalmente: la fecha de vencimiento de la factura y la disponibilidad de los cobradores por día y zona.

El principal problema es elaborar una ruta óptima que tenga en cuenta todas las variables disponibles: fecha de vencimiento de la factura, disponibilidad de los cobradores, tráfico de la zona, disponibilidad del cliente para pagar, horario del establecimiento, mercancía entregada, meteorología, etc. que mediante un proceso manual es inviable.

El objetivo principal es maximizar los cobros por visita minimizando los costes y optimizando el flujo de caja:

- Minimizar los km recorridos.
- Minimizar consumo de combustible.
- Minimizar tiempo de cobro por cliente.
- Maximizar la probabilidad de cobro.
- Mejorar la productividad de los cobradores.
- Minimizar tiempo de supervisores en definición de rutas y enfocarles en el análisis de datos.

3 Oportunidad de negocio

Este proyecto surge a partir de la necesidad detectada por el cliente de mejorar la eficiencia de la fuerza de cobradores de la compañía con dos objetivos claros:

- Garantizar que todas las ventas realizadas sean cobradas a los clientes en los plazos acordados, tanto en ventas al contado como en los pagos aplazados, asegurando la liquidez necesaria, ya que los retrasos en los pagos generan habitualmente costes financieros para poder afrontar los pagos a proveedores, nóminas y devolución préstamos.
- Optimizar las rutas de cobranza para reducir costes, identificar patrones de comportamiento de los clientes para aumentar el porcentaje de éxito en las operaciones de cobro y monitorizar la actividad de los cobradores para adaptar las rutas de forma dinámica, proponer oficinas bancarias más próximas para realizar ingresos de efectivo, etc.

OPTIMIZA desarrollará la aplicación que permitirá el análisis de los datos disponibles para obtener información del histórico de rutas, flujos de caja, impagos, importe de las operaciones, geolocalización de clientes, cobradores y vehículos, etc.

A partir de este análisis y mediante la aplicación de algoritmos de inteligencia artificial, OPTIMIZA identificará patrones de comportamiento de los clientes que permitan organizar las rutas de los cobradores teniendo en cuenta la probabilidad de realizar los cobros en una primera visita y adaptar las rutas mejorando la eficiencia de los cobradores.

OPTIMIZA acompaña a Grupo Venado en todo el proceso, desde el análisis del problema, hasta su resolución, y después analizando los resultados y buscando nuevas oportunidades de mejora en los procesos.

Las propuestas de mejora serán validadas por el cliente y se desarrollarán en colaboración con él. Es fundamental la implicación del cliente en el proyecto facilitando acceso a la información y poniendo a disposición del proyecto su infraestructura, lo que facilitará la posterior puesta en producción y la utilización del sistema por parte de los usuarios, lo que permitirá además un seguimiento de los resultados obtenidos mediante las métricas adecuadas y con el feedback de los propios usuarios para poner el valor las mejoras realizadas e identificar puntos de mejora a la solución aportada y/o analizar otros procesos susceptibles de ser mejorados.



Ilustración 4. Proceso. Fuente propia

4 Supuestos de usuario

Los usuarios que se beneficiarán del uso de la herramienta desarrollada por OPTIMIZA serán los implicados en el proceso de cobranzas:

- Planificador de rutas de cobranza: dispondrá de información necesaria para planificar las rutas de cobranza con criterios de mejora en la eficiencia de las mismas:
 - o Patrones de comportamiento del cliente: horario de apertura del establecimiento, mejor franja horaria para realizar la visita, ...
 - o Asegurar que la mercancía se ha entregado antes de incluir el establecimiento en la ruta de cobranza.

- Ayuda en la gestión de los recursos (cobradores, vehículos, etc.)
- Mejora en la gestión y dimensión de la fuerza de cobradores.
- Departamento financiero: mejorará en la gestión de los flujos de caja, tendrá información actualizada y en tiempo real de los pagos realizados y de los que están pendientes, así como información fiable sobre las previsiones de cobro. Dispondrá de información fiable sobre las facturas cobradas con su timestamp e importe correspondiente.
- Cobradores: Dispondrán de información adecuada de la ruta de cobranza a realizar, mejorando el % de éxito en cada visita y ganando en eficiencia en su trabajo diario. Podrán disponer también de información de rutas alternativas en caso de incidencias en el tráfico y propuesta de oficinas bancarias próximas para poder realizar ingresos de efectivo, mejorando por tanto la seguridad al no tener que desplazarse con grandes cantidades de dinero encima.
- Todos: Dispondrán de un módulo de reporting con informes personalizados de la actividad → cuadros de mando, autoconsumo de informes, etc.

Como beneficio colateral, afecta a la producción ya que tener un correcto flujo de caja permite realizar una planificación mejor en la compra de bienes primarios para la elaboración de productos.

5 Fuentes de datos

5.1 Encuestas

La información básica sobre los procesos existentes en el flujo de cobranzas del grupo Venado, se obtendrá a partir de encuestas realizadas en las diferentes áreas implicadas.

5.1.1 Cobranzas

Cuestionario 1	
Área	Oficina de Santa Cruz. Cobranzas
Puesto	Responsable de cobranzas (Sandra Salazar)

Pregunta	Respuesta
¿Cómo se preparan las rutas?	Programamos las rutas (circuito de mercado) (control del crédito), realizamos y detallamos a quienes ir a cobrar, por ejemplo, tanto las instituciones como los supermercados pagan en un día en concreto, entonces

Cuestionario 1

Área	Oficina de Santa Cruz. Cobranzas
Puesto	Responsable de cobranzas (Sandra Salazar)

Pregunta	Respuesta
	<p>se los tiene apartado. Y realizamos un seguimiento de los cobradores</p> <p>A día de hoy las rutas se preparan en función de los vencimientos de pago como único criterio</p>
¿Cuándo y cómo recibe el cobrador las rutas?	El cobrador recibe las rutas que va a recorrer a diario a primera hora de la mañana
¿Cada cuánto tiempo informa el cobrador del estado de cobros?	El cobrador informa del estado de los cobros una vez al mes
¿Cuántos cobradores hay en esta oficina?	En esta oficina existen 19 cobradores
¿Cómo están divididos? ¿Cuántos son chóferes y cuantos van a pie?	De éstos, 7 son cobradores a pie y 12 son chóferes cobradores
¿Qué tipo de incidencias ha habido con los cobros?	<p>Principalmente nos encontramos con que el pedido no ha llegado al cliente cuando se pasa a cobrar, creando una situación incómoda</p> <p>En alguna ocasión se ha desviado mercancía a clientes no dados de alta que se beneficiaban de los descuentos de clientes que no llegaban a recibir la mercancía.</p>
¿Qué tipo de incidencias ha habido con los cuadros de cuentas?	Desde hace ya varios años la única incidencia de este tipo viene relacionada con aquellos cobradores que han tenido que utilizar parte de los cobros para manutención y combustible.
¿Qué porcentaje de ventas son al contado?	En este momento el 70% de las ventas son al contado.
¿Cuánto tiempo se tarda en la planificación diaria y que se hace en esa planificación?	Aproximadamente una hora diaria.

Cuestionario 2

Área	Oficina de La Paz. Cobranzas
Puesto	Responsable de cobranzas (Cecilia Monroy)

Pregunta	Respuesta
¿Cómo se preparan las rutas?	Si, todos los días armamos las rutas, a quienes visitan y el monto a cobrar
¿Cuándo y cómo recibe el cobrador las rutas?	La ruta se entrega al cobrador al inicio de la jornada después de mantener la reunión diaria en la que se designan las rutas.
¿Cada cuánto tiempo informa el cobrador del estado de cobros?	El seguimiento de los cobradores se realiza diariamente sobre el trabajo diario.
¿Cuántos cobradores hay en esta oficina?	En esta oficina existen 17 cobradores
¿Cómo están divididos? ¿Cuántos son chóferes y cuantos van a pie?	De éstos, 6 son cobradores a pie y 11 son chóferes cobradores
¿Qué tipo de incidencias ha habido con los cobros?	El cliente se niega a pagar o está ausente del establecimiento en el momento de la visita del cobrador
¿Qué tipo de incidencias ha habido con los cuadros de cuentas?	Ninguno que yo recuerdo o se hubiera dado en lo que he estado trabajando.
¿Qué porcentaje de ventas son al contado?	Aproximadamente el 70% de las ventas son al contado
¿Cuánto tiempo se tarda en la planificación diaria y que se hace en esa planificación?	Una hora diaria como mínimo

5.1.2 Finanzas

Cuestionario 3	
Área	Finanzas
Puesto	Gerente Administrativo Financiero (Felipe Villarroel)

Pregunta	Respuesta
¿Cómo se preparan las rutas?	<p>Los responsables de organizar las rutas son los administradores de cada oficina.</p> <p>En la empresa actualmente tenemos BeeTrack es la herramienta de planificación con horario de pedidos y planificación de entregas del día siguiente, este sistema de planificación de los pedidos y las rutas que harán los camiones para entregar los productos es el pionero en la empresa.</p>
¿Cuándo y cómo recibe el cobrador las rutas?	Las rutas se entregan a diario a los cobradores. En el caso de ausencia de los responsables, la planificación de la ruta no es la más adecuada y aumenta el riesgo de descuadre en las cobranzas ya que la verificación no se realiza
¿Cada cuánto tiempo informa el cobrador del estado de cobros?	Depende de la oficina, puede llegar a ser mensual, hay supervisión directa en las tres regionales más grandes, Santa Cruz, La Paz y Cochabamba, pero en las otras regionales más pequeñas la supervisión recae en el jefe de la regional que tiene funciones más variadas entre las de supervisar las cobranzas es una de ellas.
¿Cuántos cobradores hay en esta oficina?	N/A
¿Cómo están divididos? ¿Cuántos son chóferes y cuantos van a pie?	N/A
¿Qué tipo de incidencias ha habido con los cobros?	N/A

Cuestionario 3

Área	Finanzas
Puesto	Gerente Administrativo Financiero (Felipe Villarroel)

Pregunta	Respuesta
¿Qué tipo de incidencias ha habido con los cuadros de cuentas?	<p>Se producen descuadres por falta de control de la actividad diaria de los cobradores</p> <p>Es necesario optimizarlas rutas y medir la productividad de los cobradores para conseguir:</p> <ul style="list-style-type: none"> • Mejorar la productividad • Identificar las mejores rutas en cuanto a tiempo de ejecución, probabilidad de cobro • Verificación de entrega de mercancía, pedidos y cobranzas • Generar estadísticas e indicadores que midan la actividad de las áreas de cobranzas • Gestionar la rotación de los cobradores • Realizar un ajustado pago de la parte variable del salario al cobrador en base a un objetivo determinado a inicio de mes, ya que actualmente este cálculo se realiza con el mes cerrado y con datos que son variables.
¿Qué porcentaje de ventas son al contado?	Aproximadamente el 70% de las ventas son al contado y el 30% a crédito
¿Cuánto tiempo se tarda en la planificación diaria y qué se hace en esa planificación?	Esta planificación se realiza desde cada oficina y requiere un tiempo que los responsables deberían dedicar a otras actividades

5.1.3 Business Intelligence

Cuestionario 1	
Área	Business Intelligence
Puesto	Gerente de Sistemas y Business Intelligence (Susana Durán)

Pregunta	Respuesta
¿Cómo se preparan las rutas?	Las rutas se preparan de manera manual por parte de la supervisión de cobranzas en base a los estados de cuentas, información generada en el sistema SGV, sólo de los supermercados se tienen días específicos para ir a cobrar.
¿Cuándo y cómo recibe el cobrador las rutas?	El cobrador debe ir a la oficina diariamente reunirse una hora con el supervisor y después sale a cobrar con la información de que ruta tomar, a que cliente visitar y sobre cuánto y cuándo debe cobrar. Es muy importante recalcar que la gran mayoría de los cobros se hacen en efectivo, el porcentaje es aproximadamente un 86%.
¿Cada cuánto tiempo informa el cobrador del estado de cobros?	El cobrador marca su información en su dispositivo, pero tiene una implicancia proactiva por su parte, está disponible de manera inmediata en sistema, pero no es consumida de esa forma por el área.
¿Cuántos cobradores hay en esta oficina?	N/A
¿Cómo están divididos? ¿Cuántos son chóferes y cuantos van a pie?	N/A
¿Qué tipo de incidencias ha habido con los cobros?	N/A
¿Qué tipo de incidencias ha habido con los cuadros de cuentas?	N/A
¿Qué porcentaje de ventas son al contado?	Aproximadamente el 70% de las ventas son al contado y el 30% a crédito

Cuestionario 1	
Área	Business Intelligence
Puesto	Gerente de Sistemas y Business Intelligence (Susana Durán)

Pregunta	Respuesta
¿Cuánto tiempo se tarda en la planificación diaria y que se hace en esa planificación?	La planificación se realiza en aproximadamente 1 hora, en base a información acumulada hasta el día en concreto.

5.2 Base de datos corporativa

Actualmente se tienen tres orígenes de datos:

- Beetrack: Es una aplicación con la que se gestionan las rutas para la entrega de mercancía, estas rutas se generan manualmente y lo que hace beetrack es el seguimiento en base a un dispositivo que va en los camiones de reparto. Alguno de los choferes que se encargan de estas entregas también realizan labores de cobrador y es en ese caso cuando envían la información al sistema central para el tracking de los cobros. Disponemos de la información recogida en una tabla de lo que envían los choferes – cobradores.

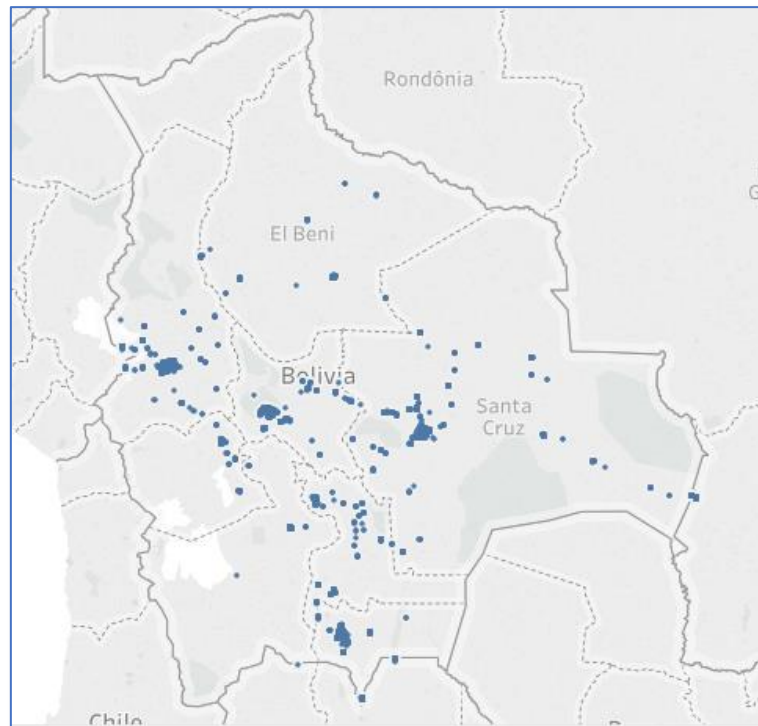


Ilustración 5. Ubicación clientes

- Base de datos de SwissBites: Base de datos que se obtiene del sistema SwissBites (SGV). Tiene un subconjunto de tablas que permiten al departamento de BI obtener la información necesaria para la generación de distintos informes.
 - Ventas/día/mes
 - Días festivos
 - Cobranzas/día/mes
- Base de datos de Cobranzas: Se recoge el “core” de las ventas/órdenes y cobros de Venado. De esta base de datos nos interesa principalmente la siguiente información:
 - Clientes: contiene la información del cliente incluyendo su geolocalización, vendedor asociado y tipo de lista de precio
 - Visitas: Almacena la información de las visitas que han realizado cada cobrador, almacenando la geolocalización del cliente, la hora de la visita, y observaciones en caso de no ser realizado el cobro.
Como dato de partida, encontramos que algo más de un 2% de las visitas realizadas por los cobradores, resultan fallidas (en una primera estimación estaríamos hablando de 1.6 M\$ anuales).

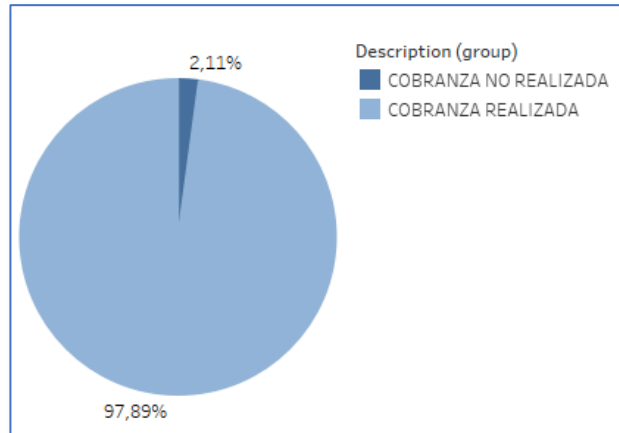


Ilustración 6. Resultado de las visitas de cobradores

- Ordenes: contiene información de la orden en relación a las visitas para la cobranza y la fecha de entrega.
 - Dispatch_order: relación de datos de que camión lo ha entregado, peso, si está o no facturado.
 - Truck: Datos de los camiones que realizan la entrega.
- Empleado: Información de los cobradores

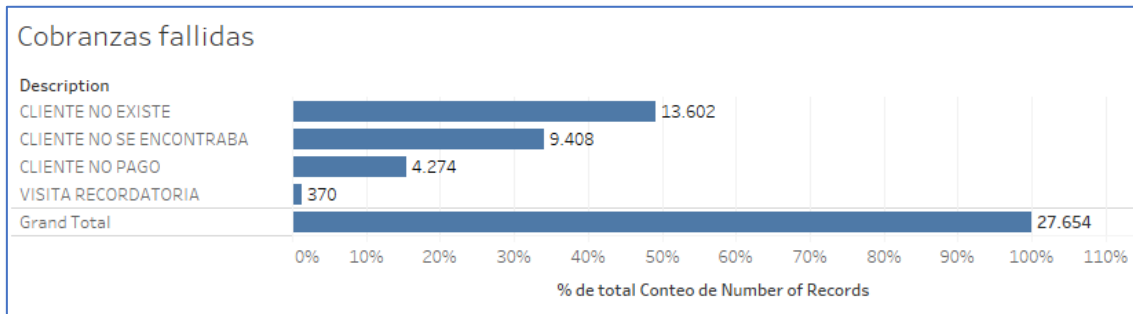


Ilustración 7. Cobranzas fallidas. Fuente: Venado

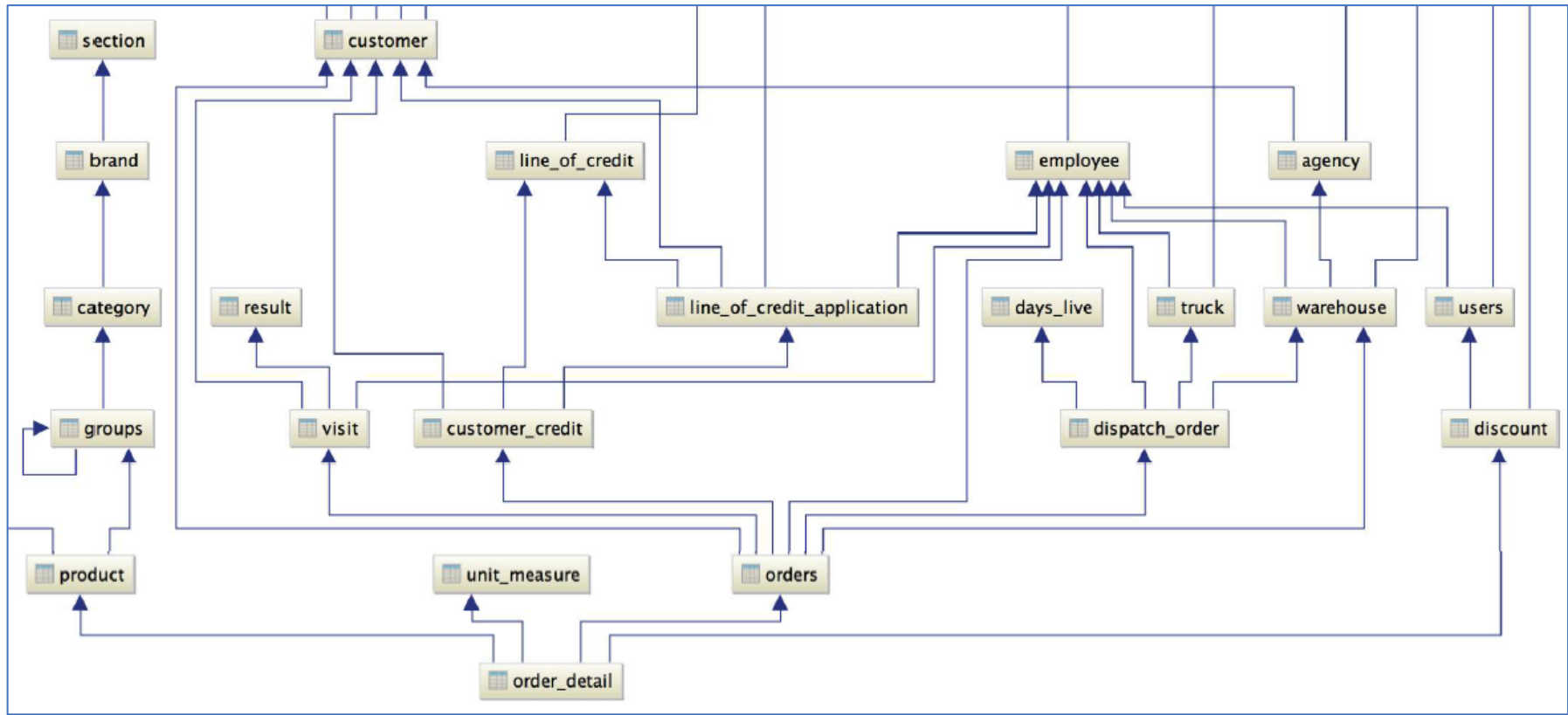


Ilustración 8. Estructura de tablas

5.3 Datos generados y de terceros

Como parte de los datos que tenemos que generar con el modelo de optimización de ruta hay información adicional que se tendrá que almacenar y afectara a varios procesos que pasamos a detallar.

Una parte importante del estudio que se hace es también pautar el comportamiento de los clientes de Venado. No todos ellos pagan en el momento de la entrega de la mercancía y depende de varios factores (algunos de ellos difícilmente medibles ya que no son medidas operativas) para definir cuándo van a pagar y por lo tanto generar el modelo de ruta mejor para el cobro.

- Seguimiento de pedidos: confirmación de entrega de mercancía -> planificación de cobranza → configuración de rutas

- Gestión de entregas pendientes: Algunas de las incidencias detectadas en las cobranzas, se debe a que no se ha realizado la entrega de la mercancía a los clientes, es necesario, por tanto, identificar las causas que provocan estas incidencias, que realmente no se deberían catalogar como impagos, sino que deberían identificarse para no incluir esos clientes en la ruta de cobranzas hasta que la mercancía no haya sido entregada. Como posibles causas de falta de entrega, podrían estimarse inicialmente las siguientes:
 - Rotura stock
 - Problema de logística
 - Incidencia en la ruta
 - No hay problema, el pedido se entregará en el plazo establecido

- Pautas de comportamiento del cliente en el pago. Esta información proviene de las visitas cuando se va a realizar el cobro y son medidas por el cobrador.
 - Cliente ausente
 - No puede
 - No quiere
 - Pago aplazado

- Datos adicionales de cliente:
 - Horarios de apertura
 - Periodos de cierre: vacaciones, inventario, etc.
 - Propuestas de mejora

- Indicadores:
 - Tiempos de entrega
 - Pagos al contado vs. Pagos aplazados
 - Impagos (totales / parciales)
 - % de éxito en cobro en 1ª visita, 2ª visita ...
 - Ranking de clientes

- Geográficos
 - Mapas <http://www.liedman.net/leaflet-routing-machine/>

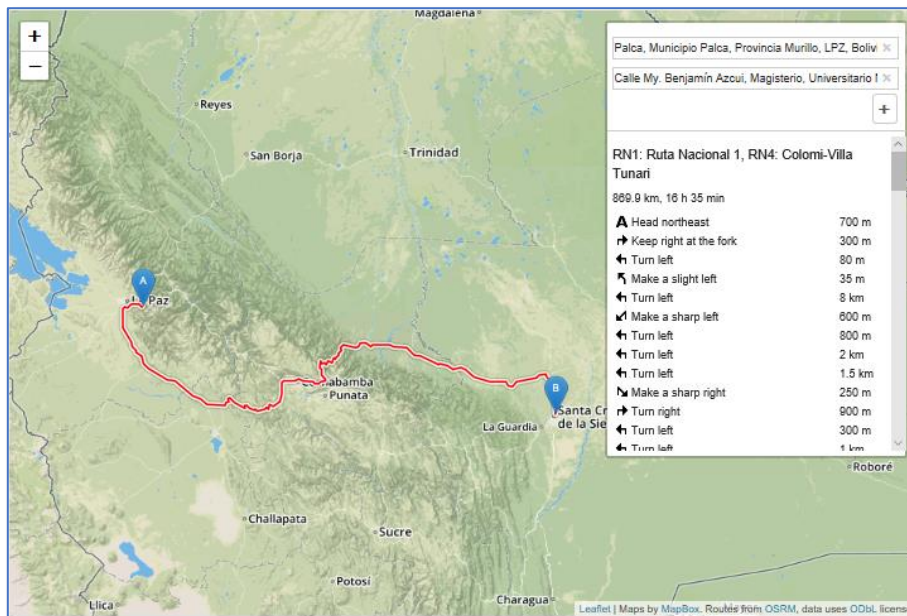


Ilustración 9. Ruta

6 Hipótesis de trabajo

A la vista de la información obtenida, y a falta de un análisis más detallado de la información disponible en las bases de datos de Venado, las hipótesis sobre las que vamos a trabajar serán las siguientes:

- El cliente necesita gestionar las cobranzas en un entorno en el que predomina el pago en efectivo y el mayor porcentaje de cobranzas se realiza al contado
- Es necesario controlar adecuadamente el proceso de cobranzas en operaciones al contado y también en operaciones a crédito ya que en este caso las cobranzas se

realizan mediante una fuerza de cobradores que visitan a los clientes, para garantizar que todos los pedidos servidos se cobran adecuadamente

- El cliente necesita información relativa a las entregas de mercancía, modalidad de pago, crédito de los clientes, ...
- Las áreas de cobranzas necesitan medir la productividad de los cobradores y mejorar el % de éxito de cobro en las visitas a clientes
- Pensamos que se puede optimizar la configuración de las rutas de cobro seguidas por los cobradores teniendo en cuenta:
 - Prioridades de los clientes
 - Horarios
 - Ubicación geográfica de los clientes
 - Localización de los cobradores
 - Estado de entrega de los pedidos
 - Histórico de operaciones
 - Probabilidad de cobro

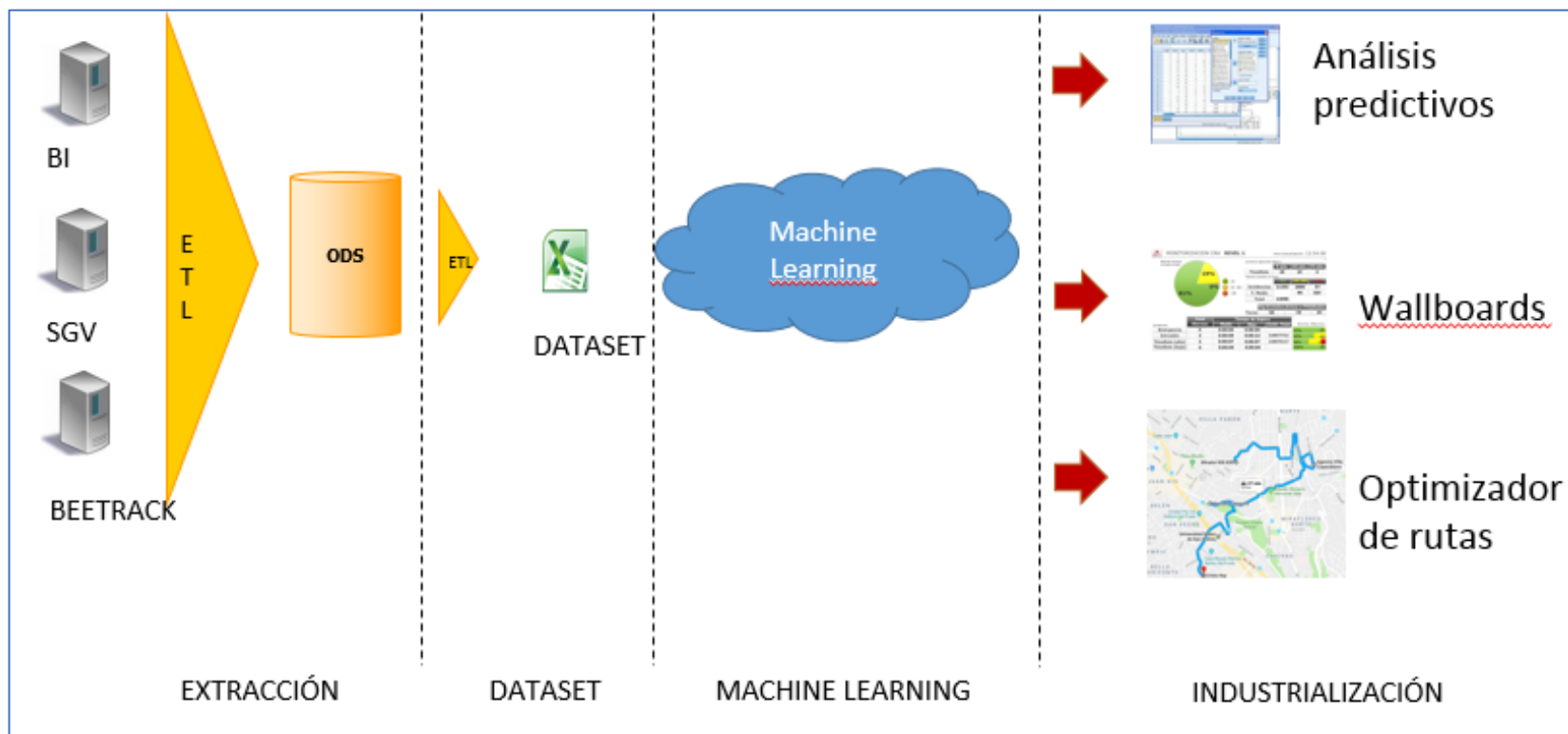


Ilustración 10. Fases del proyecto

7 Análisis y diagnóstico

En **OPTIMIZA** estamos desarrollando una solución a medida para la necesidad del grupo Venado de optimizar las rutas de cobranza que se planifican de forma manual a diario consumiendo tiempo y recursos.

No se trata únicamente de reducir ineficiencias en planificación de rutas, sino de realizar una selección automática de las visitas a realizar en función de la probabilidad de éxito en la cobranza para planificar las rutas óptimas en cada caso y que éstas estén disponibles y asignadas a cada cobrador con la mayor antelación posible.

7.1 Situación actual

El mercado ofrece diferentes soluciones para la gestión de flotas y planificación de rutas que están orientadas principalmente a cubrir necesidades de mejora de procesos de logística y transporte:

- **Simpliroute:** aplicación para la gestión de flotas que permite planificar rutas introduciendo las visitas de forma manual y facilita la gestión de vehículos y trazabilidad de mercancías mediante geolocalización y controles de entrega.

Permite la conexión con otros sistemas corporativos y facilita la gestión logística, pero requiere de la planificación manual de las rutas en la propia herramienta o la importación de estas desde otros sistemas.

- **Movertis:** al igual que la anterior es una herramienta orientada a la gestión logística, ofrece un control y monitorización de las rutas, incidencias y ofrece informes a medida de las actividades.

Se trata de una herramienta muy potente para controlar en tiempo real los vehículos, conductores y mercancías, pero la planificación de las rutas requiere de la selección manual de las visitas para optimizar la ruta en función de distancia, consumo de combustible, etc.

- **Beetrack:** es la herramienta de gestión logística que utiliza Venado para la gestión de las entregas de mercancía, como las anteriores ofrece una solución para la gestión de las rutas de entrega de mercancía y la optimización de los procesos logísticos.

La herramienta permite planificar las entregas de mercancía y la optimización de las rutas a partir de las visitas que se introducen manualmente.

Las herramientas disponibles en el mercado para la planificación y optimización de rutas están orientadas a la gestión logística y al control de flotas, sin embargo, como ya se ha indicado, la necesidad que **OPTIMIZA** pretende cubrir, es la de generar las rutas de forma automática teniendo en cuenta no solo factores geográficos para minimizar las distancias y/o los tiempos en los desplazamientos, sino los relacionados con la planificación de los cobros:

- Horarios preferentes de cobro

- Estado de entrega de la mercancía
- Fechas de vencimiento de los pagos
- Maximizar el número de visitas en cada ruta
- Minimizar el número de visitas fallidas

Es decir, el objetivo es doble, por un lado, al igual que en el caso de las aplicaciones existentes en el mercado, la reducción de costes y la optimización de tiempo y distancia recorrida, también se busca maximizar el éxito en cada visita, consiguiendo una mejora importante en el porcentaje de éxito en las cobranzas.

7.2 Debilidades, amenazas, fortalezas y oportunidades

Se trata de un proyecto a medida para el grupo Venado y por tanto contaremos con información de primera mano de las necesidades y puntos de mejora a desarrollar, será un producto escalable y que podrá crecer e incorporar nuevas funcionalidades, sin embargo, este planteamiento hace que no sea un desarrollo fácilmente paquetizable sino que requerirá un proceso de análisis completo de las necesidades de diferentes clientes.

Fortalezas.-

Fortalezas	1	2	3
Modelo que se alimenta de 2 bases de datos principales			3
Actualización en tiempo real de la ejecución de envíos			3
OPTIMIZA cuenta entre sus miembros con la directora de Business Intelligence de Venado			3
OPTIMIZA es un equipo multidisciplinar, formado por profesionales de distintas industrias con experiencias en comunicaciones, finanzas, programación y logística			3
Utilización de criterios más allá de la fecha de vencimiento y la disponibilidad de cobradores para la preparación de rutas			3
PROMEDIO			3.00

Debilidades.-

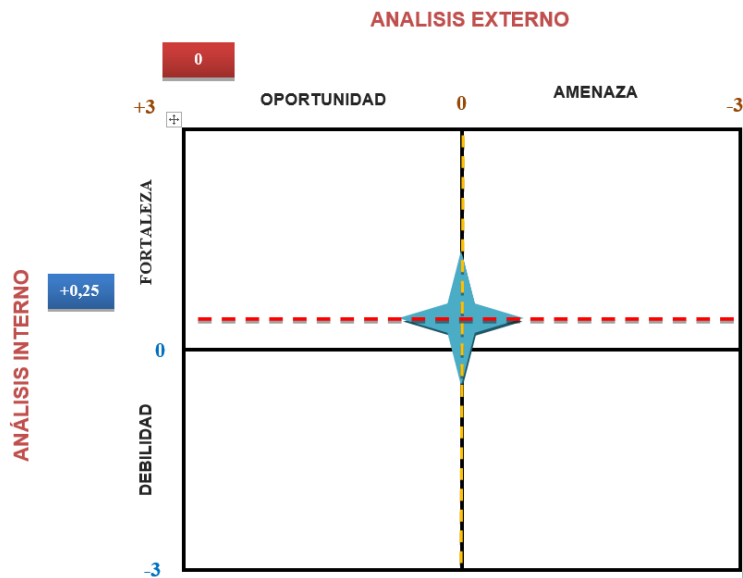
Debilidades	1	2	3
Las modificaciones de campos en la estructura las realiza una empresa externa, ralentizando el proceso		2	
Reducción de tiempo y precio de implementación, al ser un modelo implementado sobre un sistema a medida			3
PROMEDIO			2.50

Oportunidades.-

Oportunidades	1	2	3
Posibilidad de crear un perfil de comportamiento de cada cliente, optimizando de esta manera la confección de rutas		2	
Mejoras de estructuras y gestión de recursos en otras áreas y departamentos		2	
Expansión al mercado boliviano			3
PROMEDIO			2.33

Amenazas.-

Amenazas	1	2	3
Posibilidad de crear un perfil de comportamiento de cada Dependencia total del sistema de red del país.		2	
Pérdida de dispositivos móviles con los datos de los informes y posterior fuga de información-		2	
PROMEDIO			2.00



Fuente: Elaboración propia en base al análisis interno y externo

Para aplacar las debilidades:

1. Las modificaciones de campos en la estructura las realiza una empresa terciarizada y toma tiempo, se tendrá un personal del área de Sistemas que este pendiente del modelo gracias a la fortaleza del modelo de dar una optimización de recursos utilizada en el diseño de rutas óptimas para el área de cobranzas.
2. Tiempo y precio de implementación, al ser un modelo implementado sobre un sistema a medida tendrá 3 impactos muy importantes para la organización, Actualización en tiempo real de la ejecución del envío de productos y las cobranzas haciendo uso de 2 bases de datos y al final se sostendrá un modelo sobre distintas variables.

Para aplacar las Amenazas:

1. La dependencia total del sistema de red del país, el área de Sistemas de Grupo Venado está terminando de colocar fibra en todas sus distribuidoras y fábricas.
2. Pérdida de dispositivos con los reporting al descubierto y posterior fuga de información, el área de Sistemas de Grupo Venado está instalando a todos los equipos corporativos un sistema de control de megas e instalaciones de app, por lo que se podrá bloquear el equipo de manera inmediata.

8 Modelo de Negocio

El modelo de negocio que **OPTIMIZA** está desarrollando es el que se presentó en los documentos de anteproyecto en los que se identificaron y describieron los aspectos más importantes a tener en cuenta:

- **Socios clave**, entre los que se encuentran distintos departamentos del propio cliente implicados directamente en los procesos de cobranzas.
- **Actividades clave**, para definir claramente las necesidades, analizar los datos disponibles e identificar usuarios y funcionalidades de la aplicación, conexión con las fuentes de datos y con los sistemas de reporting.
- Definir la **relación con los clientes**, que en este caso serán principalmente clientes internos, los propios departamentos usuarios de la herramienta para la optimización de las operaciones, la gestión financiera, etc.
- Estructura de costes, fuentes de ingresos, recursos.

La propuesta de valor, por tanto, será cubrir una necesidad importante de identificar y priorizar las visitas a realizar en cada ruta de cobranzas atendiendo a criterios logísticos (tiempo, distancia recorrida, ...) y criterios de negocio, básicamente maximizar el importe cobrado y mejorar el flujo de caja, minimizando además las visitas fallidas.

Como consecuencia, y no menos importante, el objetivo de **OPTIMIZA** es reducir el nivel de frustración de los cobradores motivado por las visitas fallidas y la necesidad de volver a visitar al cliente.

Las áreas de actividad desarrolladas son:

- Análisis de datos
- Identificación de objetivos a visitar
- Planificación de la ruta
- Registro de resultados
- Definición de métricas y reporting,

A continuación, se describe el plan de acción y las actividades a desarrollar para conseguir los objetivos planteados.

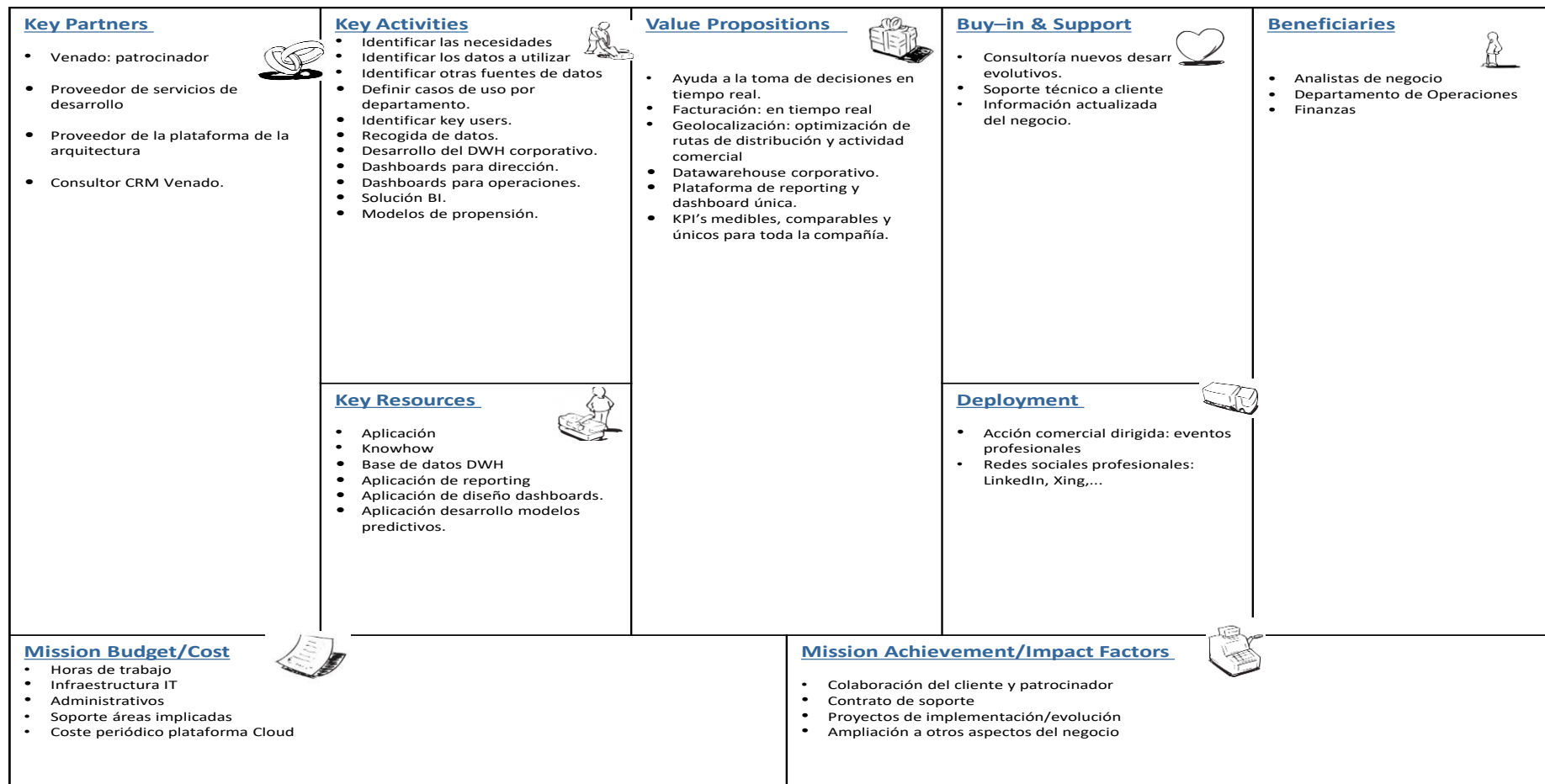


Ilustración 11. Modelo de Negocio proyecto optimización de rutas de cobranzas. Industrias Venado

El análisis de la base de datos de Industrias Venado, así como las sesiones de trabajo organizadas conjuntamente con los distintos interesados, ha permitido identificar de forma más adecuada algunos aspectos del modelo de negocio:

8.1 Key Partners

Al tratarse de un proyecto desarrollado y personalizado para Industrias Venado, éste es el principal partner y patrocinador del proyecto. Durante las fases iniciales se han realizado diferentes acciones con las áreas implicadas:

- BI: Acceso a la base de datos y varias sesiones con los especialistas para identificar los aspectos clave que resultan de interés para el proyecto
- Responsables de cobranzas: para obtener información relativa a los procesos actuales que nos ha permitido entender mejor el problema e identificar puntos de mejora
- Departamento comercial
- Dirección de la empresa para trasladar a las áreas implicadas la importancia de la mejora buscada en el rendimiento de las cobranzas y las ventajas que proporcionará a la organización esta mejora en el proceso.

8.2 Key Activities

Tras realizar diferentes entrevistas con las áreas interesadas del cliente, se han identificado áreas de actividad en las que será necesario actuar para conseguir los objetivos buscados, y que será necesario planificar adecuadamente:

- Identificar adecuadamente los clientes a incluir en cada ruta, teniendo en cuenta la información actual en base de datos:
 - o Distribuidores
 - o Cobradores
 - o Deuda pendiente y fecha de vencimiento de la misma
- Medir adecuadamente la dimensión del problema. Para ello es necesario identificar claramente qué porcentaje de visitas son fallidas y el cobrador necesita volver al cliente para poder cobrar toda o parte de la deuda pendiente. Será necesario asumir una fase inicial en la que los datos no serán del todo fiables ya que solo en los últimos meses de 2018 se registran el 100% de las visitas fallidas.
- Desarrollar la aplicación que permitirá diseñar las rutas de cobranzas a partir de los datos existentes
- Desarrollar la aplicación y DWH, así como el modelo predictivo que permita además asignar a cada cliente una prioridad relativa a la probabilidad de éxito en la cobranza que sirva para decidir si se le incluye o no en la ruta de cobranza.
- Planificar adecuadamente el despliegue de la solución en los distintos distribuidores.
- Establecer los mecanismos de monitorización y medida de indicadores que permitan evaluar los resultados del modelo y ajustarlo para mejorar los resultados.

Teniendo en cuenta que el modelo de cobranzas actual de Industrias Venado, está bastante extendido en Bolivia, desde OPTIMIZA, nos planteamos la posibilidad de ofrecer la solución a otros clientes.

Como principales competidores del grupo Industrias Venado, y posibles objetivos para dar a conocer el sistema serían:

- Nestlé
- Grupo Alicorp

8.3 Key Resources

En la fase actual del proyecto el recurso principal del que se está obteniendo información y sobre el que se está trabajando es SGV, la base de datos PostgreSQL corporativa de ventas y cobranzas de Industrias Venado.

Adicionalmente y como recursos clave para el desarrollo del proyecto, OPTIMIZA cuenta con un entorno desarrollo en la nube de Amazon para las aplicaciones que formarán parte de la solución:

- ETL conectado a la base de datos SGV para ingesta y pre-procesado de datos
- DWH para el almacenamiento de los datos que alimentarán el modelo de previsión y la aplicación de cálculo de rutas.
- APP: Selección de clientes por distribuidor y cobrador para incluir en la planificación de ruta
- OSRM: Cálculo de rutas
- Tableau: Seguimiento y reporting...

9 Plan de acción

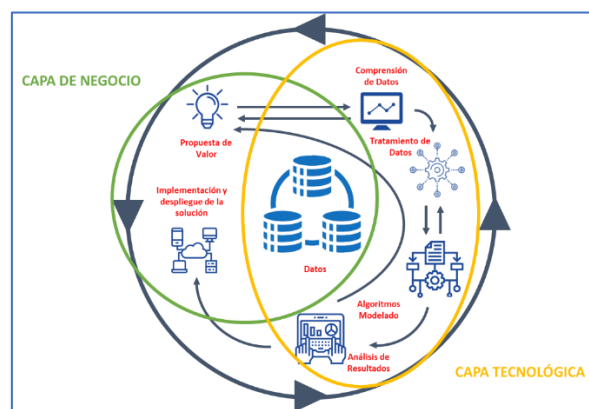


Ilustración 12. Plan de acción

9.1 Alcance, objetivos y métricas

Alcance: generación, asignación y seguimiento automático de las rutas de visitas para los distintos cobradores de las 5 delegaciones.

Objetivos:

- Optimizar el tiempo de cobro por ruta, reduciendo el número de visitas de los cobradores a un cliente, mediante el estudio de sus patrones de conducta. (sacar métrica de base de datos de la media actual y la mejora que se espera). Esto a su vez, incrementará el dinero recogido por ruta y se podrá incrementar el número de clientes a visitar por ruta.
- Reducir la dependencia de los supervisores que son los que actualmente generan estas rutas. Basados en su experiencia y conocimiento, son ellos los que diseñan y al día siguiente evalúan las rutas de cobro
- Poder tener una mejor previsión diaria de flujo de caja que afectaría a otros departamentos para realizar sus operaciones al mejorar el ratio de cobranzas exitosas en la primera visita. Se dispondrá de información diaria frente al análisis semanal actual.

Para la consecución de estos objetivos necesitaremos desarrollar las siguientes actividades:

- Extracción de la geo-localización de los clientes y posicionamiento en un mapa de los mismos.
- Cruce con datos de órdenes de entrega: Ayudará a no realizar visitas a clientes que no han recibido el pedido todavía.
- Cruce con ventas: Puede ser una variable más a la hora de priorizar los cobros saber pautas y cantidad en las ventas.
- Datos de clientes: Datos objetivos de métodos de pago de los clientes. (fechas de cobro)
- Obtención de datos históricos de comportamiento de clientes en el pago: tanto de rutas preestablecidas como de información acerca de nuestros clientes en cuanto a qué hora es la mejor para la recogida del cobro.

El sistema a desarrollar deberá implantarse paulatinamente en las cinco delegaciones, siendo La Paz la primera y la que se tomará a modo de entrenamiento para posteriormente implantar el sistema en el resto. Se da un periodo de 6 meses para que el sistema esté plenamente funcionando en todas las delegaciones y medir si el número de repetición de las visitas se ha reducido.

Los principales involucrados son:

- Departamento de Cobranzas: Los supervisores tendrán un doble trabajo en el período de implantación al tener que evaluar si las rutas propuestas por el sistema han sido más eficientes y ha mejorado el número de visitas y reducido el tiempo que los cobradores

pasan en cada cliente, para dar feedback a OPTIMIZA y que éste realice los ajustes necesarios.

- Departamento Financiero: Debe revisar si lo ingresado mejora con respecto al anterior sistema.
- Departamento de Informática: dedicados a la implantación del sistema.
- OPTIMIZA: deberá monitorizar (junto con los supervisores e informática) los resultados y ajustar en caso de requerirlo los algoritmos para ajustarse al resultado esperado.

9.1.1 Ratio de visitas duplicadas por cliente

Actualmente el ratio de revisitas al cliente es excesivo (50,85% en 2018) y la tendencia según el siguiente gráfico es alcista.



Ilustración 13. Histórico de revisitas

El objetivo es llegar a un índice por debajo del 30 % en el primer año y reducir este índice hasta llegar a un 15-10% en el segundo año.

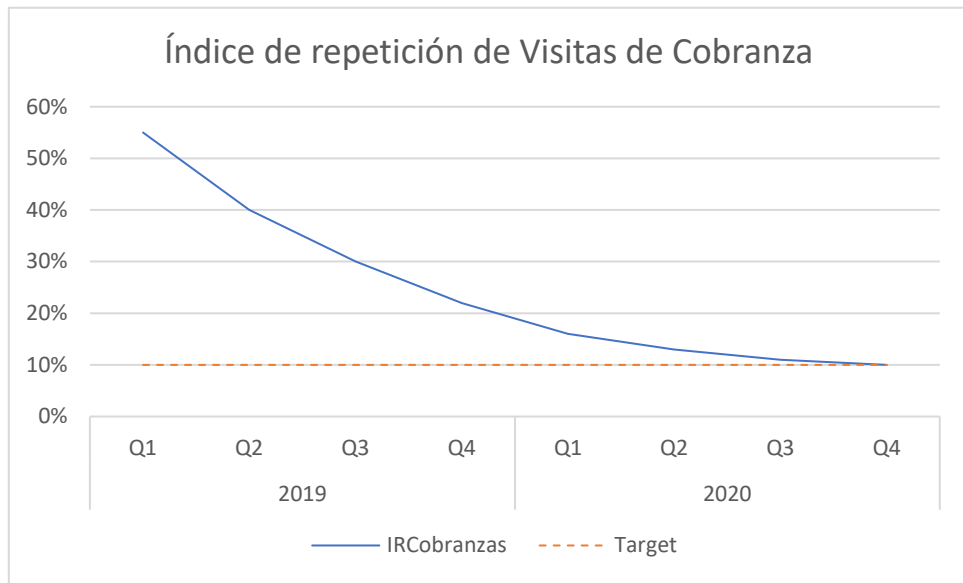


Ilustración 14. Evolución del índice de repetición de visitas

Mejora del ratio de éxito de cobranza en la primera visita e incremento del flujo de caja por ruta en la primera visita.

En la siguiente gráfica se muestra cobro total por mes.

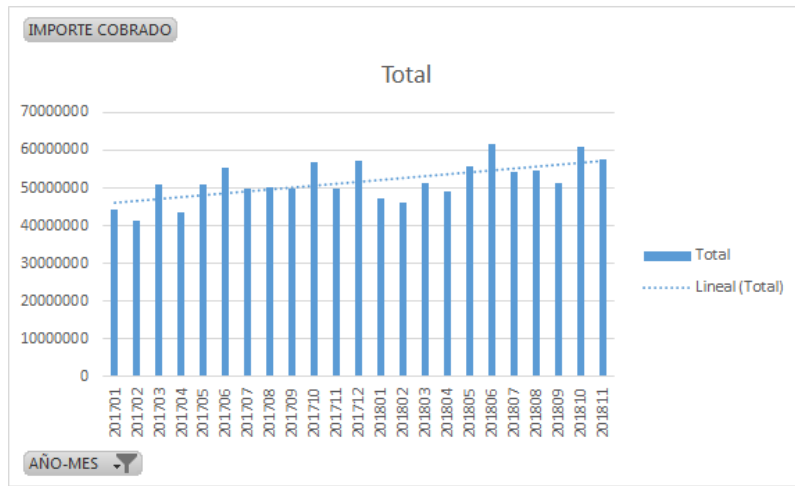


Ilustración 15. Histórico de cobranzas

Aunque la tendencia es alcista se observan picos de baja productividad que se atenuarán con la inclusión de esta solución.

El objetivo es incrementar el flujo de caja en la primera visita, es decir, mejorar la eficiencia en los cobros que repercutirá en un ahorro de costes al reducir el número de visitas, en una mayor eficiencia en el trabajo de los cobradores y por tanto en su desempeño y satisfacción.

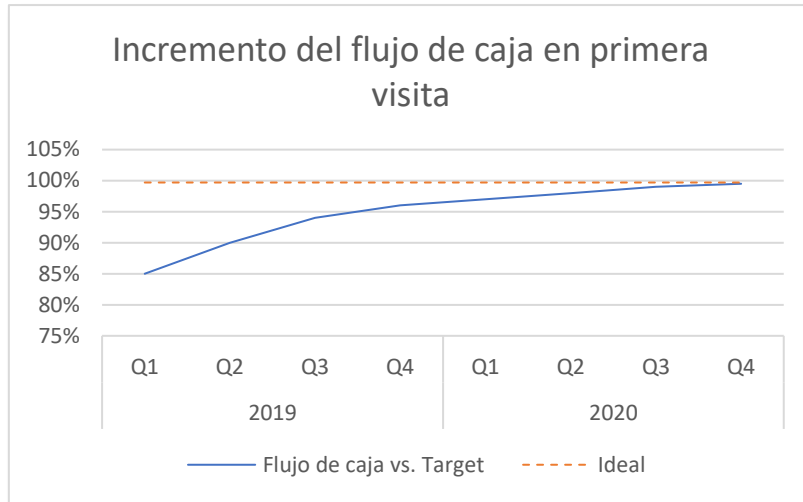


Ilustración 16. Evolución del flujo de caja

% de cobros fallidos frente al total.

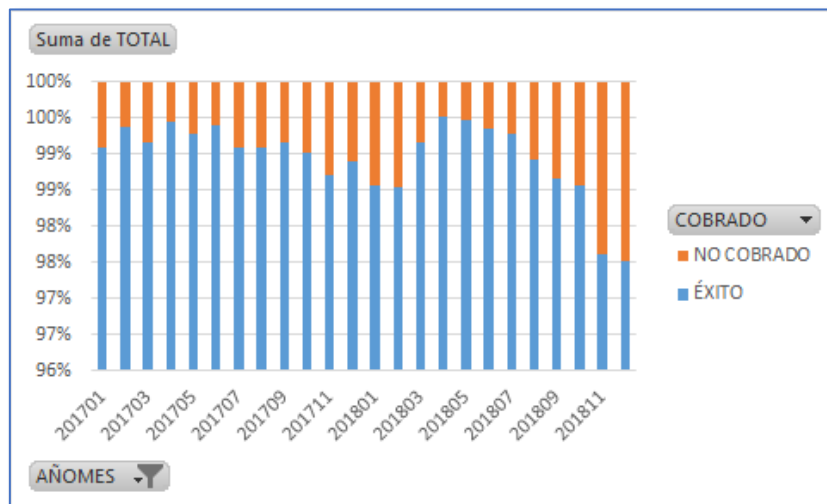


Ilustración 17. % de Éxito en las cobranzas

El ratio de éxito en cobranzas está bajando a razón de 1 punto cada dos meses. El incremento de clientes y un índice de impago mayor hace que sea necesario incluir herramientas que ayuden a optimizar las rutas de cobradores.

9.1.2 Reducción del tiempo dedicado a la planificación de rutas

Uno de los factores que provocan ineficiencia en el proceso de cobranzas es la planificación de las rutas que, actualmente se realiza de forma manual y diariamente.

En esta tarea están implicados tanto los responsables de cobranzas de cada uno de los distribuidores, que dedican en torno a una hora diaria a la planificación, como los cobradores, que no pueden empezar su trabajo hasta que está finalizada la planificación.

Para mejorar esta eficiencia, la propuesta de OPTIMIZA ofrecerá una planificación automática de rutas que el cobrador podrá recibir a primera hora, o incluso el día anterior, con lo que se liberará a los responsables de la tarea de planificar rutas y les permitirá centrarse en tareas de control y gestión para evaluar los resultados.

Con este sistema los cobradores podrán comenzar su trabajo a primera hora de la mañana, conseguirán una ruta con mayor probabilidad de éxito en las visitas a realizar, y mejorarán por tanto su desempeño.

Adicionalmente se analizará la posibilidad de aumentar el número de visitas por ruta y obtener así una mejora adicional en el rendimiento de los cobradores.

En una primera estimación, teniendo en cuenta el número de cobradores por distribuidor, y estimando una pérdida de rendimiento de una hora diaria para la planificación de las rutas, el número de horas dedicadas a esta actividad y que podrían ahorrarse serán:

Recurso	Actividad	Empleados	Horas Año
Planificador	Planificación de ruta	12	3120
Cobrador	Espera x ruta - Distribuidor		
	DIST. BENI	3	780
	DIST. COCHABAMBA	7	1820
	DIST. DISALTO	7	1820
	DIST. DISCRUZ	20	5200
	DIST. DISPAZ	8	2080
	DIST. DISTAR	11	2860
	DIST. IVSA	1	260
	DIST. ORURO	3	780
	DIST. POTOSI	2	520
	DIST. SUCRE	3	780
Horas de trabajo recuperadas			20020

Lo que supone una ineficiencia de aproximadamente un 14,4%.

El objetivo perseguido para la mejora en este proceso, partiendo de los datos de 2018 y en una estimación a dos años, sería:

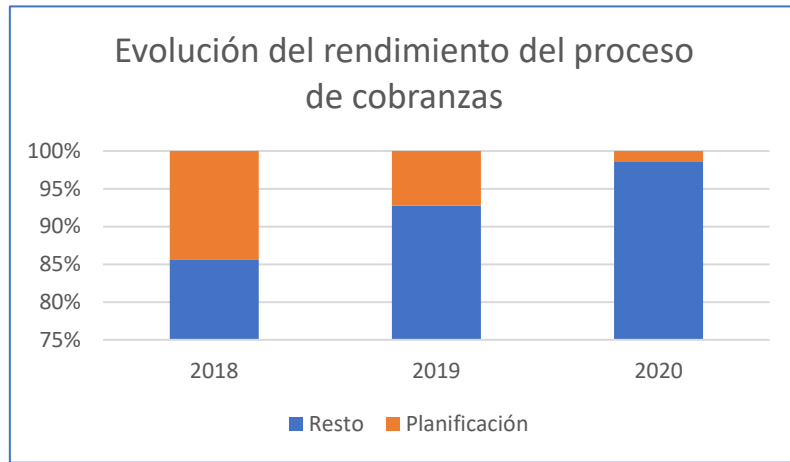


Ilustración 18. Objetivo de mejora en el rendimiento de las cobranzas

9.2 Actividades: Mapa de procesos y solución tecnológica

9.2.1 Mapa de procesos



Ilustración 19. Plan de acción

9.2.2 Solución tecnológica

Modelo Lógico

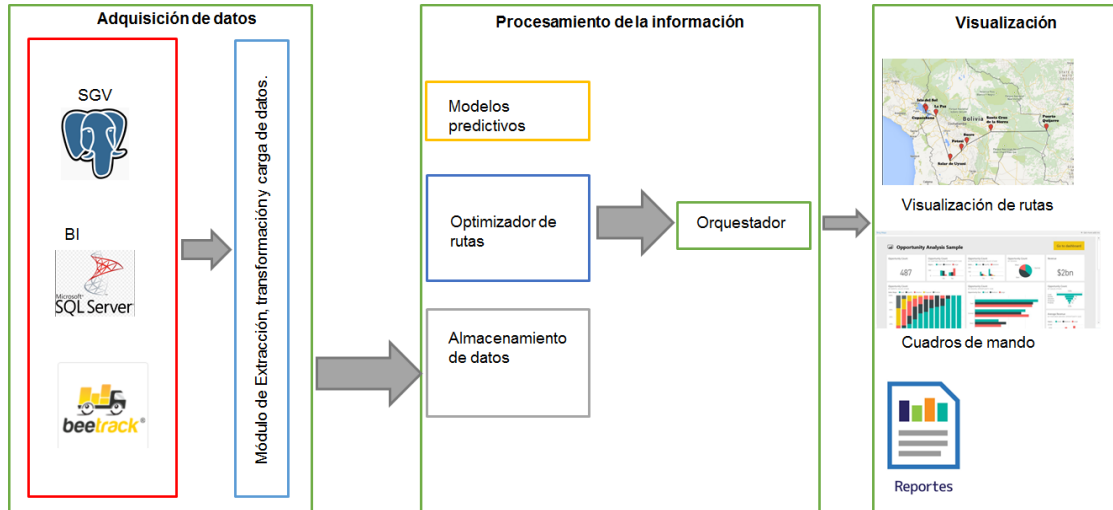


Ilustración 20. Modelo lógico

1. Identificación de los datos

La información se obtendrá de los diferentes repositorios internos de Venado:

BEETRACK	Estructurado	SQL SERVER	Estado de las visitas realizadas por los chóferes cobradores.
SGV	Estructurado	POSTGRES	Visitas realizadas por los cobradores junto con la geoposición de los clientes.
BI	Estructurado	SQL SERVER	Modelo de datos informacional para realizar los diferentes informes corporativos de Venado (Ventas, ingresos, facturación)

2. Extracción, transformación y carga de datos

Proceso para unificar las diferentes fuentes de datos en un único repositorio común, y realizar limpieza de los datos erróneos o incompletos. Se estudiará la posibilidad de crear alertas para mostrar al usuario si los datos no tienen la calidad suficiente.

3. Almacenamiento de los datos

Repositorio independiente donde se almacenarán los datasets utilizados por el modelo predictivo y los resultados de las diferentes ejecuciones y entrenamientos.

4. Validación de los datos

Mecanismo para garantizar la calidad, la integridad y la autenticidad de los datos.

5. Visualización

Formato: Se representará el mapa de la zona de Bolivia que correspondiente incluyendo la ruta asignada al cobrador. Los analistas serán autosuficientes a la hora de representar los datos a través de la herramienta corporativa.

Tipo de aplicación: Aplicación web.

Dispositivos: Móviles y ordenador.

9.2.3 Arquitectura técnica

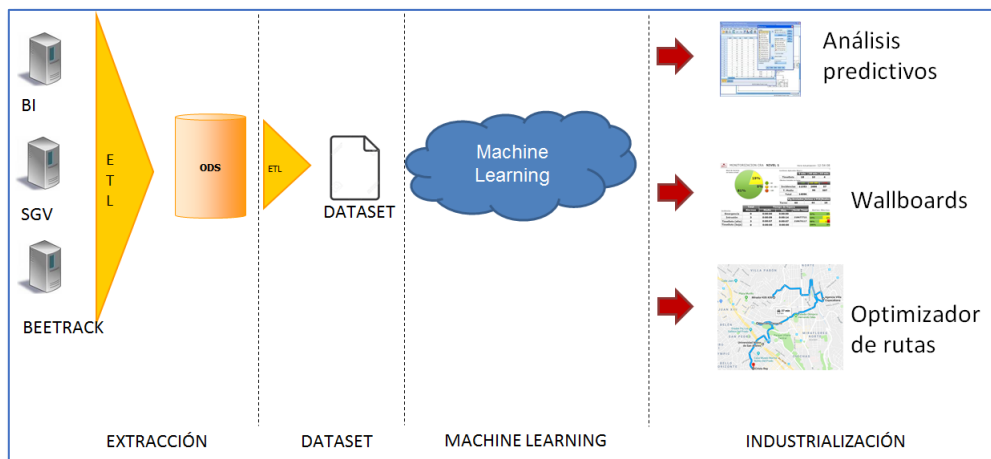


Ilustración 21. Solución tecnológica

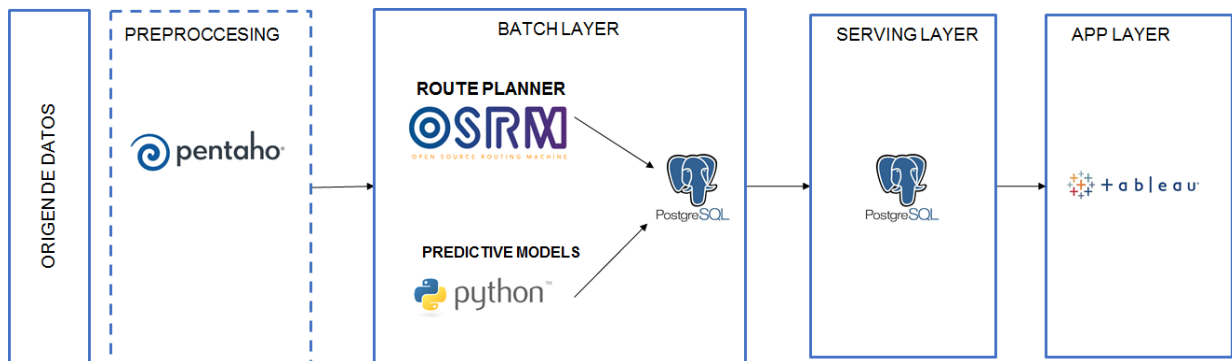


Ilustración 22. Detalle de la solución

Preprocessing: Capa encargada de conectarse a todos los orígenes de datos y cargarlos en un único destino.

En esta fase se limpiarán todos los campos, verificando tanto sus valores como si el formato es el adecuado.

Batch layer: Esta capa se divide en dos componentes: Route Planner y predictive models.

Route Planner: Se encarga de calcular las rutas óptimas para cada cobrador teniendo en cuenta los resultados del modelo predictivo correspondiente. Se utilizará la herramienta OSRM para su cálculo y su visualización.

Predictive models: Creación del modelo predictivo que ayudará al Route Planner a ser más eficiente. Se tendrán en cuenta diferentes variables como: horario de apertura, mercancía recibida, tramo horario donde se paga más habitualmente, etc...

Todos los resultados de la capa batch se almacenarán en una instancia de postgresql para su uso en tiempo de proceso.

Serving Layer: Capa de servicio basada en una nueva instancia postgresql donde se aíslan los datos que serán consultados por los analistas sin impactar en el sistema operacional.

App Layer: La capa de visualización utilizará la herramienta corporativa de Venado, Tableau, para realizar los diferentes gráficos y mapas necesarios.

9.2.3.1.1 Dataset del modelo

El dataset se compone de todas las visitas de Venado realizadas por los cobradores en los últimos n años.

```

SELECT visit.id as visit_id,
visit.visit_date,
to_char(visit.visit_date,'d') as dayWeekVisita,
to_char(visit.visit_date,'w') as weekMonthVisita,
to_char(visit.visit_date,'wy') as weekYearVisita,
visit.employee_id,
visit.customer_id,
visit.result_id,
visit.latitude,
visit.longitude,
visit.spent_time,
visit.registration_date,
visit.operation_type,

```

Ilustración 23. Detalle SQL. Variables

```

into tmp_venado_visitas
from visit inner join
  venado_cobradores as cobradores on visit.employee_id = cobradores.employee_id inner join
  customer on visit.customer_id = customer.id left outer join
  customer_credit on visit.customer_id = customer_credit.customer_id and customer_credit.status = 0 left outer join
  line_of_credit on customer_credit.line_of_credit_id = line_of_credit.id
where visit.operation_type in (1, 4) and
  visit.origin = 1 and
  customer.is_prodiverse = false and
  visit.visit_date::date between '2017-01-01' and '2019-02-28';

```

Ilustración 24. Detalle SQL. Creación de tabla temporal

Además de la información relativa a la visita (Fecha, estado de visita, cobrador,...) se complementa con diferentes indicadores muy significativos para priorizar las visitas en el modelo supervisado:

- Total Visitas satisfactorias en los últimos 5 días, 10 días, 20 días y 30 días.
- Total Visitas no satisfactorias en los últimos 5 días, 10 días, 20 días y 30 días.
- Total deuda acumulada.
- Antigüedad máxima de deuda.

9.2.3.1.2 Modelo Supervisado

Mediante la aplicación de técnicas de machine learning se pretende obtener una lista de clientes a visitar cada día priorizada en función de la probabilidad de éxito en el cobro, así como una propuesta de mejor horario para realizar dicha visita.

Entrenamiento

Partiendo del dataset descrito en el punto anterior, que contiene el histórico de visitas a clientes, así como la información más relevante de los mismos, se aplican dos algoritmos:

- **Clasificador Random forest:** La variable 'operation_type' indica el resultado de la visita (toma valor 4 cuando el cobro se ha realizado con éxito y 1 cuando la visita no ha tenido éxito y el cobro no se ha podido realizar). Se aplicará un modelo supervisado para el entrenamiento del modelo, dividiendo el dataset en dos subconjuntos de datos, uno para **entrenamiento (65%** de los datos) y otro para **test (35%** restante). Después de aplicar el modelo a los datos de entrenamiento, se evalúa con los de test. Puesto que tenemos los datos reales de éxito (modelo supervisado), los comparamos con los obtenidos de la predicción realizada con los datos de test mediante una matriz de confusión, que nos permite ver unos resultados bastante aceptables del modelo.

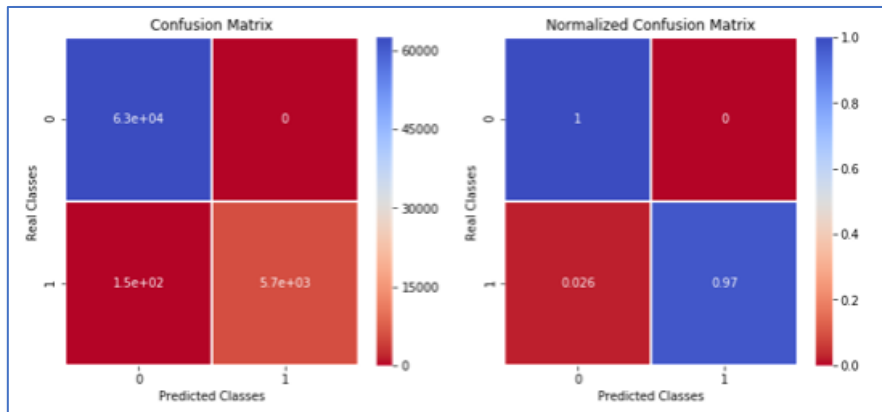


Ilustración 25. Evaluación del modelo de clasificador Random-forest

```

---Classification Report---
              precision    recall  f1-score   support

0               1.00         1.00         1.00         62556
1               1.00         0.97         0.99         5868

 micro avg       1.00         1.00         1.00         68424
 macro avg       1.00         0.99         0.99         68424
 weighted avg    1.00         1.00         1.00         68424
    
```

Una vez clasificadas las visitas como OK o KO, obtenemos la probabilidad de que efectivamente ese sea el resultado.

- **Regresión Random forest:** Para la estimación de la mejor hora para realizar la visita, se aplica un algoritmo de regresión en el que el resultado buscado es la hora a la que se realiza la visita.
Para el entrenamiento del modelo se usarán los datos del dataframe filtrando únicamente las visitas OK
Aplicando mínimos cuadrados para evaluar el modelo, se observa que el resultado es aceptable:

```
MSE 0.18767027284311674  
MAE 1.6442404949305196  
RMSE 0.4332092714186952
```

Predicción

Una vez entrenado el modelo, los datos de entrada que se usarán para realizar la predicción de clientes a visitar, se realizará en base a las variables utilizadas en el entrenamiento para aquellos clientes que tienen una deuda que vence el día para el que se realizará la planificación.

Para ello la aplicación recuperará los datos relativos a la deuda con fecha de vencimiento igual a la fecha para la que se realizará la planificación, así como el resto de variables necesarias para el estudio.

Este nuevo dataframe se usará para aplicar los algoritmos predictivos del modelo de clasificación para estimar qué deuda tiene mayor probabilidad de ser cobrada, y al modelo de regresión para estimar la mejor hora para realizar la visita del cobrador.

Output

Una vez que se han obtenido los valores de probabilidad de cobro y mejor hora para realizar la visita, se añaden al dataset de deuda y se almacenan en una nueva tabla en la base de datos que será la tabla de entrada para la aplicación de planificación de rutas.

Como prueba de concepto, para incluir o no al cliente en la planificación de visitas, se establece como criterio que la probabilidad de OK en el cobro esté por encima del 85%

A medida que los datos de histórico de visitas se vayan incrementando, será posible realizar un entrenamiento más completo del modelo y seguir evaluando la relevancia de las distintas variables, así como ir introduciendo mayor número de éstas, de forma que la capacidad de estimar la probabilidad de cobro mejore.

En una fase posterior, será posible analizar los clientes con probabilidades de cobro más bajas, para definir acciones que mejoren el proceso por ejemplo, modificando la línea de crédito, cambiando el día o la hora de las visitas, o realizando acciones comerciales específicas.

9.2.3.1.3 Planificador de rutas y dashboards

La aplicación a desarrollar se encarga de mostrar las rutas optimizadas por cobrador.

Existen 10 centros de distribución en todo Bolivia y un número variable de cobradores por cada centro de distribución. Los cobradores tienen un grupo de clientes asignado y son los que se ocupan de sus cobros.

Actualmente todos los cobradores se dirigen al centro de distribución y allí los supervisores son los encargados de obtener la lista de clientes que se deben visitar basado en el vencimiento de deuda que corresponda a ese día. Una vez chequeado la información se imprime y se reparte con los cobradores, que salen a las rutas y realizan el cobro apuntando las coordenadas del cliente, la cantidad cobrada en caso de que la visita haya sido exitosa y la razón por la que no ha sido fallida en caso contrario.

Proceso

- Se obtienen los datos de los clientes, datos de vencimiento, historial de pago y número de visitas fallidas y exitosas para realizar el estudio del patrón de conducta con respecto al pago.

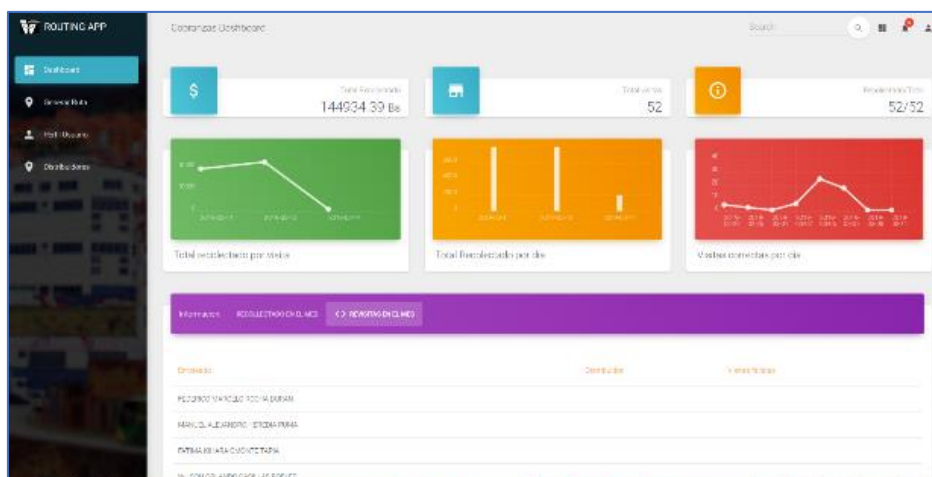


Ilustración 26. Dashboard

- Se entrena el modelo y se obtienen los datos para el día.
- Se almacenan los datos de probabilidad de cobro en la tabla junto con la información de los clientes para realizar el cálculo de las rutas.
- La aplicación web obtiene los datos de vencimiento de pago de cliente y calcula las rutas llamando al trip service.
- Se muestran las rutas en la web, siendo modificables por el supervisor.

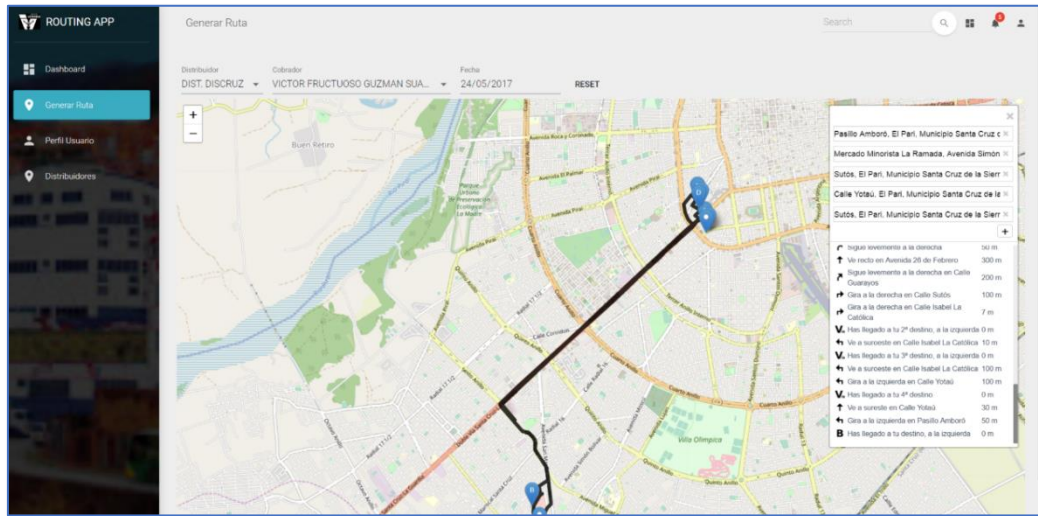


Ilustración 27. Planificación de rutas

- Se guardan en la web para que los cobradores puedan consultarlos.

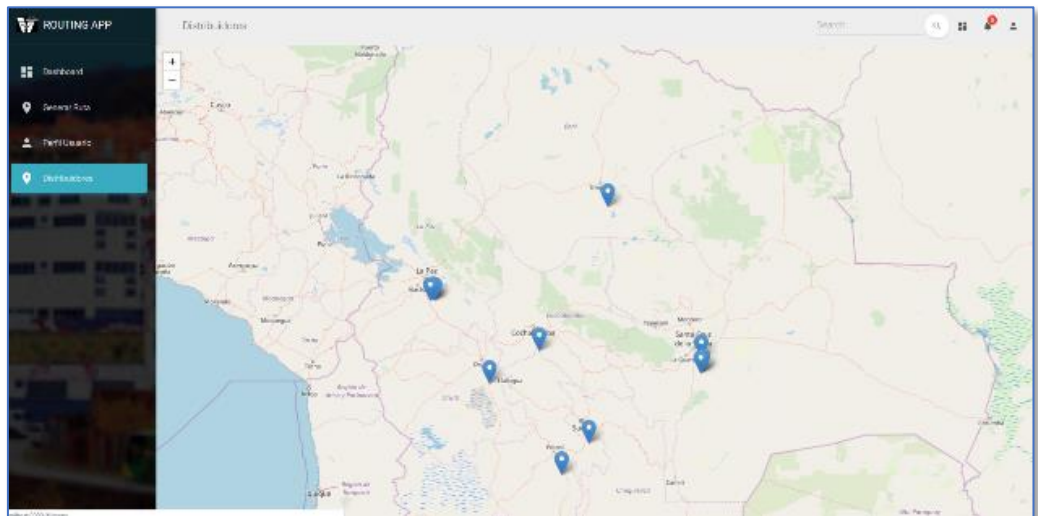


Ilustración 28. Organización por distribuidores

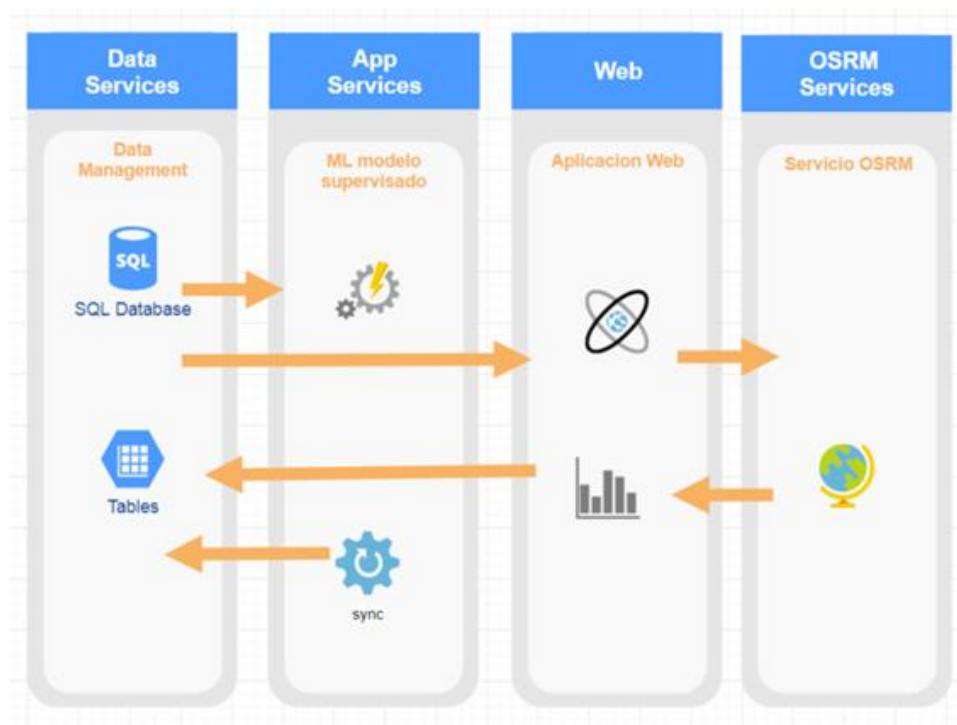


Ilustración 29. Diagrama de Bloques de la aplicación

La aplicación tendrá una parte web y una parte móvil (sólo cobradores) y la funcionalidad se dividirá dependiendo de los roles.

Roles

- Desde el punto de vista del administrador
 - Gestionar los usuarios y los permisos de la aplicación.
 - Gestionar las rutas de configuración de los distintos componentes de la aplicación
 - Chequear las notificaciones de la aplicación
- Desde el punto de vista del supervisor
 - Dashboard, con la información relevante acerca de las visitas y lo recolectado por los cobradores.
 - Generador de rutas: se selecciona distribuidor y cobrador, junto con la fecha y generar la ruta óptima para el cobro.
- Desde el punto de vista del cobrador
 - Ver la ruta con la información de los clientes a visitar. En una segunda fase se añadirá el feed de las visitas que ahora mismo se controla desde otro aplicativo.

Tecnología

App - frontend – aplicación realizada con REACT

App - backend – node para la comunicación con

OSRM API – servicio desplegado para la consulta de

BD – PostgreSQL

9.3 Organización y recursos

OPTIMIZA ofrece servicios avanzados para ayudar a las empresas en sus procesos de evolución tecnológica.

El equipo está formado por profesionales con amplia experiencia en diferentes áreas trabajando con el objetivo de satisfacer las necesidades de los clientes aplicando metodologías de gestión ágil de proyectos.



Ilustración 30. Organización

9.4 Gestión del tiempo

El proyecto de optimización de rutas de cobranza se plantea en tres fases:

9.4.1 Definición del modelo de negocio y planificación

En esta fase se desarrollarán los hitos de definición del proyecto:

- Definición del problema.
- Investigación y toma de datos.

- Análisis, modelo y plan de acción.

9.4.2 Desarrollo

- Diseño de la aplicación y desarrollo.
 - o POC.
 - o Optimización y pruebas.
- Planificación del despliegue.

9.4.3 Despliegue

- First Office Application (FOA): Primera implementación en distribuidor de Santa Cruz, en esta implementación se optimizará el proceso de instalación, configuración y pruebas del sistema y se elaborará y ejecutará el plan de pruebas que garantice el correcto funcionamiento del sistema.

Desarrollo de entorno de monitorización y reporting: medida de indicadores de servicio y optimización de procesos.

- Despliegue en el resto de distribuidores, siguiendo los procesos optimizados en la etapa de FOA.
 - o Implementación
 - o Pruebas
 - o Monitorización y seguimiento
 - o Optimización

9.4.4 Entrenamiento continuo y aprendizaje

Teniendo en cuenta que los datos disponibles en lo relativo al histórico de intentos fallidos de cobranza y repetición de visitas, nos dan únicamente información fiable de los últimos meses, será necesario, una vez creado el modelo de previsión, establecer un periodo de entrenamiento y aprendizaje del modelo para conseguir mejorar los resultados y conseguir cumplir los objetivos de manera eficaz.

9.4.5 Desarrollo de negocio

Una vez iniciado el despliegue del sistema, se dispondrá de una versión estable en producción, por tanto, en paralelo con el proyecto de implantación en Venado, se iniciarán los procesos de desarrollo de negocio para buscar nuevos clientes.

- Identificación de posibles clientes: prospección de empresas del sector que tengan necesidades similares a Venado en lo relativo a la optimización de los procesos de cobranzas.
- Definición de la estrategia de ventas.
 - Plan de marketing: visibilidad del caso de éxito en eventos profesionales.

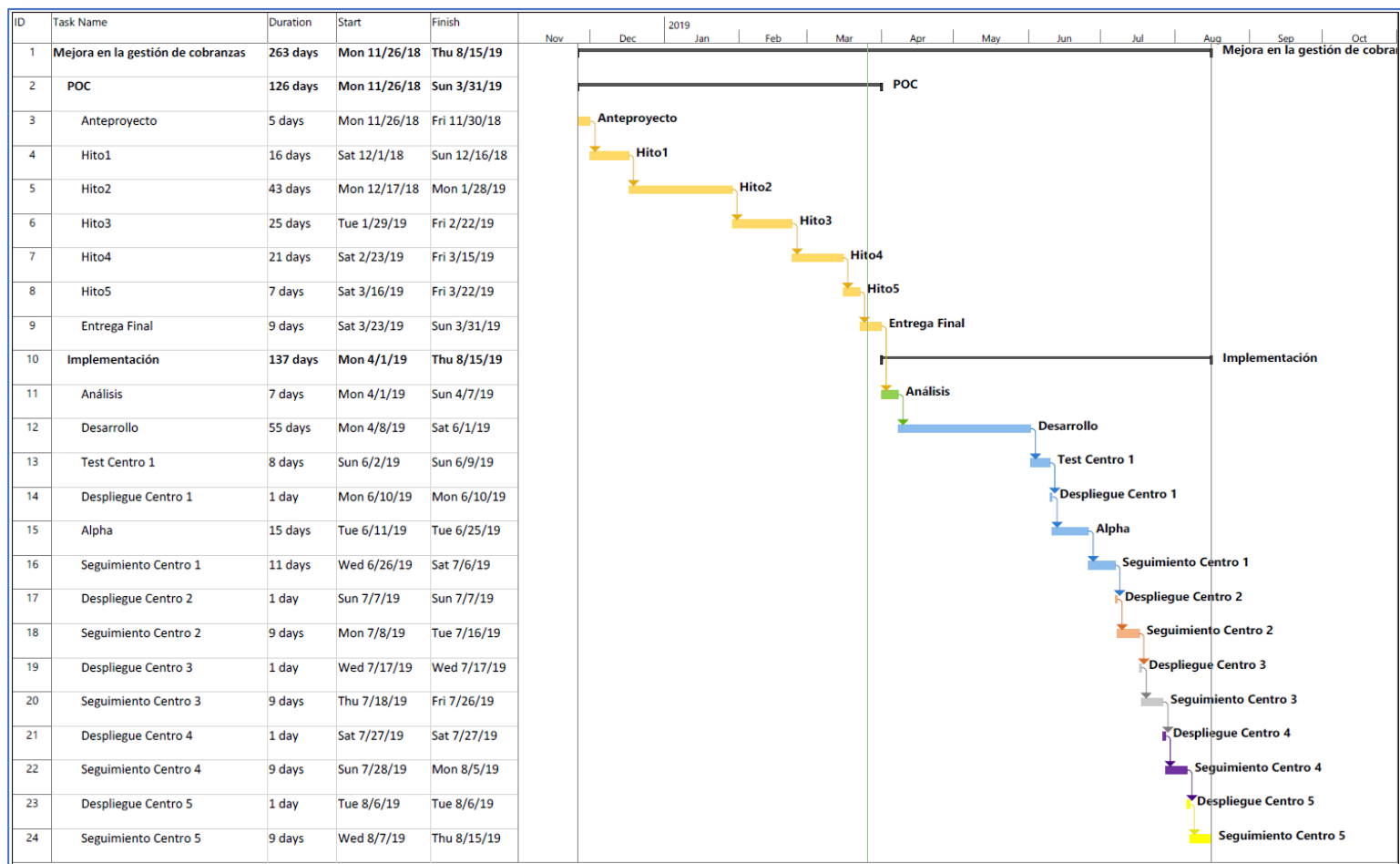


Ilustración 31. Planificación a alto nivel

10 Proyecto de creación

10.1 Plan de inversión

El plan de inversión planteado para el proyecto se divide en 2 etapas; el anteproyecto ejecutado en el tiempo previsto para el trabajo final de Máster y la implementación del mismo, una vez se cuente con la aprobación de Industrias Venado S.A.

Cuadro #1: Detalle de Tiempos del proyecto Total

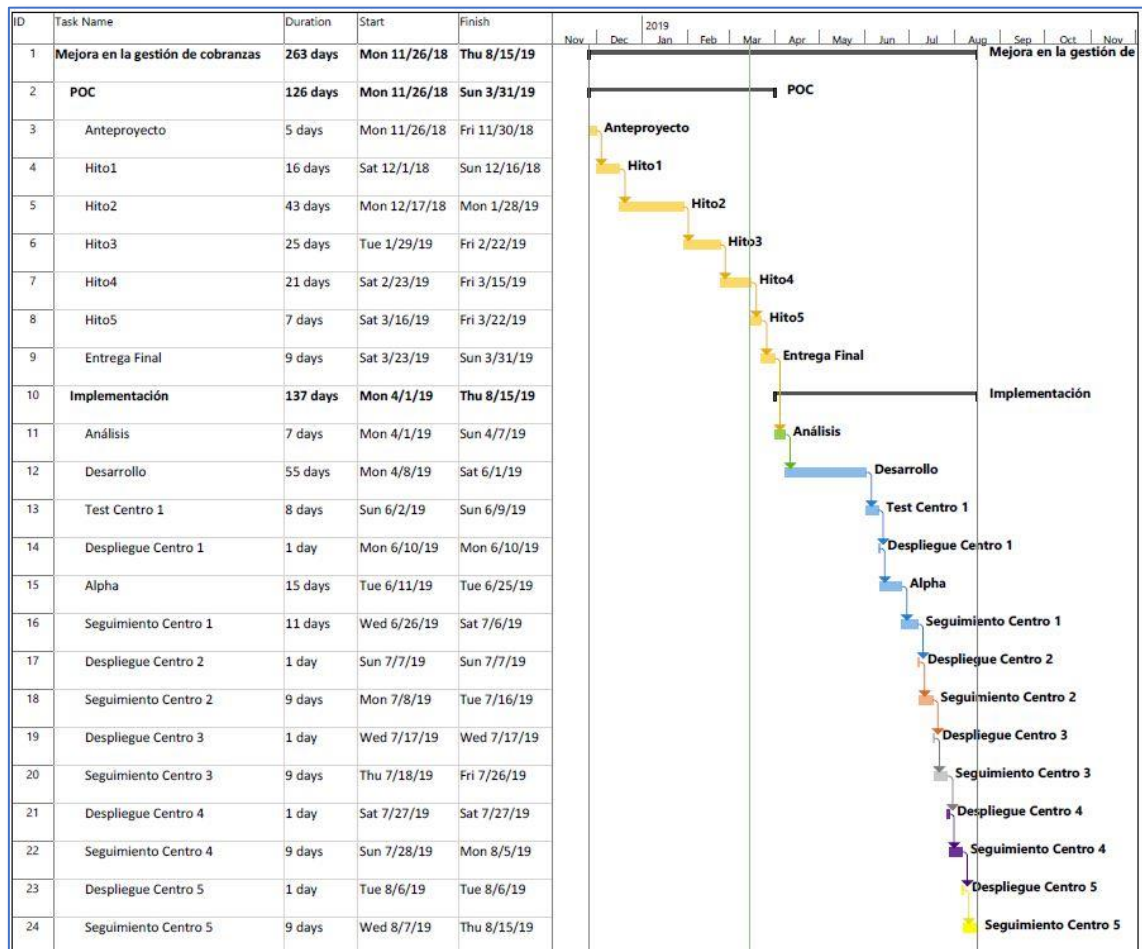


Ilustración 32. Planificación. Fuente: Elaboración Propia

10.1.1 CÁLCULO DE COSTES

	37%			67%	Margen	18%		25%
	SALARIO MENSUAL	42%	25%			Venta		Venta
		% DE CARGAS SOCIALES	% CARGOS FIJOS	TOTAL MENSUAL	COSTE Día	Día Anteproyecto	Día Implementación	
JEFE DE PROYECTO	€ 4,166.67	€ 1,750.00	€ 1,041.67	€ 6,958.33	€ 231.94	€ 273.69	€ 289.93	
ANALISTA	€ 2,916.67	€ 1,225.00	€ 729.17	€ 4,870.83	€ 162.36	€ 191.59	€ 202.95	
DESARROLLADOR	€ 2,166.67	€ 910.00	€ 541.67	€ 3,618.33	€ 120.61	€ 142.32	€ 150.76	
TEST	€ 1,833.33	€ 770.00	€ 458.33	€ 3,061.67	€ 102.06	€ 120.43	€ 127.57	
SOPORTE	€ 1,583.33	€ 665.00	€ 395.83	€ 2,644.17	€ 88.14	€ 104.00	€ 110.17	

Fuente: Elaboración Propia

10.1.2 POC

La prueba de concepto realizada tomando información de un departamento de Bolivia y como detalle para el Máster de Business Intelligence y Big Data en la EOI se realizará en 126 días. La prueba de concepto tiene un Costo Total de 37.120 €, a nivel personal requerido para su desarrollo.

Tabla #1: Presupuesto detallado de los recursos del POC

Resource Name	Initials	Std. Rate	Accrue At	Base Calendar	Cost
Jefe de Proyecto Anteproyecto	J	€273.69/day	Prorated	Optimiza	€ 4,828
Analista Anteproyecto	A	€191.59/day	Prorated	Optimiza	€ 10,380
Desarrollador Anteproyecto	D	€142.32/day	Prorated	Optimiza	€ 17,932
Test Anteproyecto	T	€120.43/day	Prorated	Optimiza	€ 2,407
Soporte Anteproyecto	S	€104.00/day	Prorated	Optimiza	€ 1,572

€ 37,120

Fuente: Elaboración Propia

10.1.3 Fase de implementación

Una vez aprobada el anteproyecto, los siguientes 5 meses tratarían el despliegue en las siguientes ciudades o centros, sus correspondientes pruebas y seguimientos con los recursos detallados dando un total de 42.647 €.

Tabla #2: Presupuesto detallado de los recursos de la Fase de Implementación

Resource Name	Initials	Std. Rate	Accrue At	Base Calendar	Cost
Jefe de Proyecto Implementación	J	€289.93/day	Prorated	Optimiza	€ 13,108
Analista Implementación	A	€202.95/day	Prorated	Optimiza	€ 7,284
Desarrollador Implementación	D	€150.76/day	Prorated	Optimiza	€ 19,599
Test Implementación	T	€127.57/day	Prorated	Optimiza	€ 1,021
Soporte Implementación	S	€110.17/day	Prorated	Optimiza	€ 1,636

€ 42,647

10.1.4 Proyecto Fase de soporte

Si bien se concluye el proyecto se tiene establecido realizar un trabajo de soporte por 3 años. Proveyendo en promedio 15 horas mensuales.

Tabla #3: Detalle de Tiempos de la Fase de Implementación

Año	Horas Mes	Std. Rate	Base Calendar	Cost
1	15	€52/hora	Optimiza	€ 9,360
2	15	€52/hora	Optimiza	€ 9,360
3	15	€52/hora	Optimiza	€ 9,360
				€ 28,080

Fuente: Elaboración Propia

11 Proyecto de optimización

11.1 Beneficios tangibles

11.1.1 Reducción de costes

En base a la investigación y entrevistas realizadas se ha logrado determinar que una hora diaria es tomada para realizar la planificación de ruta, haciendo así que los cobradores se movilicen hasta las oficinas de cada ciudad solo para recibir la ruta del día.

Por lo tanto, en base a los sueldos promedios y las 20 horas utilizadas mensualmente para la determinación de las rutas la optimización es entre un 12% y un 87,5% con respecto a un sueldo promedio.

Tabla #4: Detalle del valor de la optimización de la Planificación de Rutas para Cobradores

Cobradores	Cantidad de empleados	Sueldo Asumido empresa con Beneficios	Sueldo Promedio	Horas mes	Horas utilizadas en planificación de rutas	Sueldo por Hora	Optimización promedio	Optimización mensual en Euros	% de optimización de un sueldo promedio
Adm CBBA	3	€ 4,339	€ 1,446	160	20	€ 9	€ 181	€ 542	37.5%
Adm Disalto	6	€ 8,888	€ 1,481	160	20	€ 9	€ 185	€ 1,111	75.0%
Adm Discruz	7	€ 10,145	€ 1,449	160	20	€ 9	€ 181	€ 1,268	87.5%
Adm Dispaz	4	€ 6,145	€ 1,536	160	20	€ 10	€ 192	€ 768	50.0%
Adm Distar	2	€ 2,780	€ 1,390	160	20	€ 9	€ 174	€ 347	25.0%
Admon BNI	1	€ 1,386	€ 1,386	160	20	€ 9	€ 173	€ 173	12.5%
Admon ORU	2	€ 2,081	€ 1,041	160	20	€ 7	€ 130	€ 260	25.0%
Admon PSI	1	€ 1,055	€ 1,055	160	20	€ 7	€ 132	€ 132	12.5%
Admon SUC	2	€ 2,110	€ 1,055	160	20	€ 7	€ 132	€ 264	25.0%
Total general	28	€ 38,930	€ 1,390	160				€ 4,866	

Fuente: Elaboración Propia

Realizamos el mismo análisis para los supervisores y determinamos la optimización por ciudad es de 12,5% excepto La Paz, donde la optimización es del 25%.

Tabla #4: Detalle del valor de la optimización de la Planificación de Rutas para Supervisores

Supervisores	Cuenta de Empleado	Sueldo Asumido empresa con Beneficios	Sueldo Promedio	Horas mes	Horas utilizadas en planificación de rutas	Sueldo por Hora	Ahorro por hora	Optimización mensual en Euros	% de optimización de un sueldo promedio
Adm CBBA	1	€ 2,099	€ 2,099	160	20	€ 13	€ 262	€ 262	12.5%
Adm Discruz	1	€ 2,241	€ 2,241	160	20	€ 14	€ 280	€ 280	12.5%
Adm Dispaz	2	€ 4,221	€ 2,110	160	20	€ 13	€ 264	€ 528	25.0%
Adm Distar	1	€ 1,402	€ 1,402	160	20	€ 9	€ 175	€ 175	12.5%
Admon BNI	1	€ 1,325	€ 1,325	160	20	€ 8	€ 166	€ 166	12.5%
Admon ORU	1	€ 1,325	€ 1,325	160	20	€ 8	€ 166	€ 166	12.5%
Admon PSI	1	€ 1,325	€ 1,325	160	20	€ 8	€ 166	€ 166	12.5%
Admon SUC	1	€ 1,325	€ 1,325	160	20	€ 8	€ 166	€ 166	12.5%
Total general	9	€ 15,264	€ 1,696	160				€ 1,908	

Fuente: Elaboración Propia

Por lo que la propuesta con respecto a los cobradores es disminuir 1 en CBBA, 1 en Disalto, 1 en Discruz y 1 supervisor en Dispaz.

Con la propuesta realizada el beneficio tangible mensual sería el siguiente:

Tabla #5: Valor de la reestructuración de la fuerza de Cobranzas por contar con un Generador y Optimizador automático de Rutas

Tipo de empleado	Localidad	Sueldo Promedio	Cantidad	Total Optimizado
COBRADORES	Adm CBBA	€ 1,446	€ 1	€ 1,446
COBRADORES	Adm Disalto	€ 1,481	€ 1	€ 1,481
COBRADORES	Adm Discruz	€ 1,449	€ 1	€ 1,449
SUPERVISORES	Adm Dispaz	€ 2,110	€ 1	€ 2,110
	COBRANZAS			€ 6,487

Fuente: Elaboración Propia

11.2 Beneficios intangibles

El beneficio intangible de la optimización de ruta se verá contrastado en mayor productividad de la fuerza de ventas, en la reducción de la hora utilizada en la planificación de rutas, el detalle de la optimización mensual después de la reestructuración sería de 3.370 € mensuales.

Tabla #6: Valor intangible por la productividad de 1 hora de personal que no se reajusta

Tipo de empleado	Optimización mensual en Euros
COBRADORES	€ 1,945
SUPERVISORES	€ 1,215
Total general	€ 3,159

Fuente: Elaboración Propia

11.3 Análisis económico-financiero para el Cliente

Contemplando el plan de inversión necesario más el soporte por los siguientes 3 años, la reestructuración del área de cobranzas por generación y optimización automáticas de las rutas de cobranzas, a la cual se le está aplicando un 5% de crecimiento anual debido a que el incremento salarial es mandatorio en Bolivia.

La tasa de oportunidad está delimitada en 5,7%, en base a las tasas del sistema financiero boliviano.

Tabla #7: Flujo de Caja a 3 años del Proyecto

	Año 0	Año 1	Año 2	Año 3
Reestructuración de la fuerza		€ 77,847	€ 81,739	€ 85,826
Inversión en recursos de personal	€ 79,767			
Inversión en tecnología	€ 10,000			
Soporte anual		€ 9,360	€ 9,360	€ 9,360
Total Flujo de Caja	€ (89,767)	€ 68,487	€ 72,379	€ 76,466
VAN	€ 98,921			
TIR	60.0%			
PRI	1 año y 3 meses			

Fuente: Elaboración Propia

En los 3 años proyectados el Valor Actual Neto del proyecto es de 98.921 €, al día de hoy, por lo que al ser positivo al descontar los flujos a la tasa de descuento de 5,7% estamos frente a un proyecto rentable.

La Tasa interna de retorno es del 60% mayor a la tasa de oportunidad, esta quiere decir que el proyecto ofrece una tasa de 60%, que es mayor a la tasa de descuento del 5,7% utilizada para descontar los flujos.

12 Flujo de Caja OPTIMIZA

OPTIMIZA tiene unos términos de cobro a mes vencido. Para no tener Cash Flow negativo se ha necesitado hacer una inversión inicial de 7.100 € recuperable en el mes de junio.

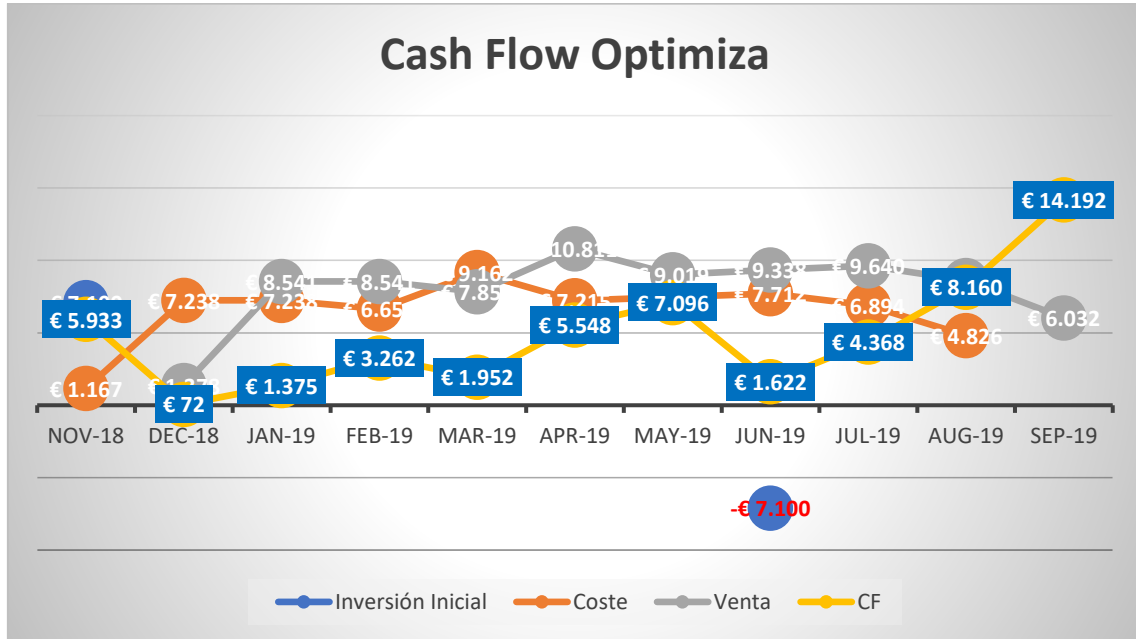


Ilustración 33. Evolución de flujos de caja

	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Sep-19
Inversión Inicial	€ 7,100							-€ 7,100			
Coste	€ 1,167	€ 7,238	€ 7,238	€ 6,654	€ 9,162	€ 7,215	€ 7,471	€ 7,712	€ 6,894	€ 4,826	
Venta		€ 1,378	€ 8,541	€ 8,541	€ 7,851	€ 10,811	€ 9,019	€ 9,338	€ 9,640	€ 8,618	€ 6,032
CF	€ 5,933	€ 72	€ 1,375	€ 3,262	€ 1,952	€ 5,548	€ 7,096	€ 1,622	€ 4,368	€ 8,160	€ 14,192

Fuente: Elaboración Propia

13 Programa detallado de recursos (en precio de venta)

Task Name	Work	Duration	Start	Finish	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Total
Cumulative					€ 1,378	€ 9,918	€ 18,459	€ 26,310	€ 37,120	€ 46,139	€ 55,477	€ 65,117	€ 73,735	€ 79,767	
Mejora en la gestión de cobranzas	3,749.43 hrs	263 days	Mon 11/26/18	Thu 8/15/19	€ 1,378	€ 8,541	€ 8,541	€ 7,851	€ 10,811	€ 9,019	€ 9,338	€ 9,640	€ 8,618	€ 6,032	€ 79,767
POC	1,863.43 hrs	126 days	Mon 11/26/18	Sun 3/31/19	€ 1,378	€ 8,541	€ 8,541	€ 7,851	€ 10,811						€ 37,120
Anteproyecto	67.6 hrs	5 days	Mon 11/26/18	Fri 11/30/18	€ 1,378										€ 1,378
Jefe de Proyecto Anteproyecto	5.6 hrs		Mon 11/26/18	Fri 11/30/18	€ 192										€ 192
Analista Anteproyecto	17.2 hrs		Mon 11/26/18	Fri 11/30/18	€ 412										€ 412
Desarrollador Anteproyecto	40 hrs		Mon 11/26/18	Fri 11/30/18	€ 712										€ 712
Soporte Anteproyecto	4.8 hrs		Mon 11/26/18	Fri 11/30/18	€ 62										€ 62
Hito1	216.32 hrs	16 days	Sat 12/1/18	Sun 12/16/18		€ 4,408									€ 4,408
Jefe de Proyecto Anteproyecto	17.92 hrs		Sat 12/1/18	Sun 12/16/18		€ 613									€ 613
Analista Anteproyecto	55.03 hrs		Sat 12/1/18	Sun 12/16/18		€ 1,318									€ 1,318
Desarrollador Anteproyecto	128 hrs		Sat 12/1/18	Sun 12/16/18		€ 2,277									€ 2,277
Soporte Anteproyecto	15.37 hrs		Sat 12/1/18	Sun 12/16/18		€ 200									€ 200
Hito2	581.37 hrs	43 days	Mon 12/17/18	Mon 1/28/19		€ 4,133	€ 7,714								€ 11,847
Jefe de Proyecto Anteproyecto	48.17 hrs		Mon 12/17/18	Mon 1/28/19		€ 575	€ 1,073								€ 1,648
Analista Anteproyecto	147.92 hrs		Mon 12/17/18	Mon 1/28/19		€ 1,236	€ 2,307								€ 3,543
Desarrollador Anteproyecto	344 hrs		Mon 12/17/18	Mon 1/28/19		€ 2,135	€ 3,985								€ 6,120
Soporte Anteproyecto	41.28 hrs		Mon 12/17/18	Mon 1/28/19		€ 187	€ 349								€ 537
Hito3	338 hrs	25 days	Tue 1/29/19	Fri 2/22/19			€ 827	€ 6,061							€ 6,888
Jefe de Proyecto Anteproyecto	28 hrs		Tue 1/29/19	Fri 2/22/19			€ 115	€ 843							€ 958

Task Name	Work	Duration	Start	Finish	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Total
Analista Anteproyecto	86 hrs		Tue 1/29/19	Fri 2/22/19			€ 247	€ 1,812							€ 2,060
Desarrollador Anteproyecto	200 hrs		Tue 1/29/19	Fri 2/22/19			€ 427	€ 3,131							€ 3,558
Soporte Anteproyecto	24 hrs		Tue 1/29/19	Fri 2/22/19			€ 37	€ 275							€ 312
Hito4	315.83 hrs	21 days	Sat 2/23/19	Fri 3/15/19				€ 1,790	€ 4,476						€ 6,266
Jefe de Proyecto Anteproyecto	23.52 hrs		Sat 2/23/19	Fri 3/15/19				€ 230	€ 575						€ 805
Analista Anteproyecto	72.23 hrs		Sat 2/23/19	Fri 3/15/19				€ 494	€ 1,236						€ 1,730
Desarrollador Anteproyecto	168 hrs		Sat 2/23/19	Fri 3/15/19				€ 854	€ 2,135						€ 2,989
Test Anteproyecto	31.92 hrs		Sat 2/23/19	Fri 3/15/19				€ 137	€ 343						€ 481
Soporte Anteproyecto	20.17 hrs		Sat 2/23/19	Fri 3/15/19				€ 75	€ 187						€ 262
Hito5	150.63 hrs	7 days	Sat 3/16/19	Fri 3/22/19					€ 2,772						€ 2,772
Jefe de Proyecto Anteproyecto	7.83 hrs		Sat 3/16/19	Fri 3/22/19					€ 268						€ 268
Analista Anteproyecto	24.08 hrs		Sat 3/16/19	Fri 3/22/19					€ 577						€ 577
Desarrollador Anteproyecto	56 hrs		Sat 3/16/19	Fri 3/22/19					€ 996						€ 996
Test Anteproyecto	56 hrs		Sat 3/16/19	Fri 3/22/19					€ 843						€ 843
Soporte Anteproyecto	6.72 hrs		Sat 3/16/19	Fri 3/22/19					€ 87						€ 87
Entrega Final	193.68 hrs	9 days	Sat 3/23/19	Sun 3/31/19					€ 3,563						€ 3,563
Jefe de Proyecto Anteproyecto	10.08 hrs		Sat 3/23/19	Sun 3/31/19					€ 345						€ 345
Analista Anteproyecto	30.97 hrs		Sat 3/23/19	Sun 3/31/19					€ 741						€ 741
Desarrollador Anteproyecto	72 hrs		Sat 3/23/19	Sun 3/31/19					€ 1,281						€ 1,281
Test Anteproyecto	72 hrs		Sat 3/23/19	Sun 3/31/19					€ 1,084						€ 1,084
Soporte Anteproyecto	8.63 hrs		Sat 3/23/19	Sun 3/31/19					€ 112						€ 112
Implementación	1,886 hrs	137 days	Mon 4/1/19	Thu 8/15/19						€ 9,019	€ 9,338	€ 9,640	€ 8,618	€ 6,032	€ 42,647

Task Name	Work	Duration	Start	Finish	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Total
Análisis	74.48 hrs	7 days	Mon 4/1/19	Sun 4/7/19						€ 2,090					€ 2,090
Jefe de Proyecto Implementación	18.48 hrs		Mon 4/1/19	Sun 4/7/19						€ 670					€ 670
Desarrollador Implementación	56 hrs		Mon 4/1/19	Sun 4/7/19						€ 1,421					€ 1,421
Desarrollo	704 hrs	55 days	Mon 4/8/19	Sat 6/1/19						€ 6,928	€ 9,338	€ 301			€ 16,568
Jefe de Proyecto Implementación	145.2 hrs		Mon 4/8/19	Sat 6/1/19						€ 2,201	€ 2,966	€ 96			€ 5,262
Analista Implementación	118.8 hrs		Mon 4/8/19	Sat 6/1/19						€ 1,260	€ 1,699	€ 55			€ 3,014
Test Implementación	440 hrs		Mon 4/8/19	Sat 6/1/19						€ 3,467	€ 4,674	€ 151			€ 8,292
Test Centro 1	189.43 hrs	8 days	Sun 6/2/19	Sun 6/9/19								€ 4,015			€ 4,015
Jefe de Proyecto Implementación	21.12 hrs		Sun 6/2/19	Sun 6/9/19								€ 765			€ 765
Analista Implementación	40.32 hrs		Sun 6/2/19	Sun 6/9/19								€ 1,023			€ 1,023
Desarrollador Implementación	64 hrs		Sun 6/2/19	Sun 6/9/19								€ 1,206			€ 1,206
Test Implementación	64 hrs		Sun 6/2/19	Sun 6/9/19								€ 1,021			€ 1,021
Despliegue Centro 1	10.63 hrs	1 day	Mon 6/10/19	Mon 6/10/19								€ 246			€ 246
Jefe de Proyecto Implementación	2.63 hrs		Mon 6/10/19	Mon 6/10/19								€ 96			€ 96
Test Implementación	8 hrs		Mon 6/10/19	Mon 6/10/19								€ 151			€ 151
Alpha	159.6 hrs	15 days	Tue 6/11/19	Tue 6/25/19								€ 3,697			€ 3,697
Jefe de Proyecto Implementación	39.6 hrs		Tue 6/11/19	Tue 6/25/19								€ 1,435			€ 1,435
Test Implementación	120 hrs		Tue 6/11/19	Tue 6/25/19								€ 2,261			€ 2,261
Seguimiento Centro 1	140.8 hrs	11 days	Wed 6/26/19	Sat 7/6/19								€ 1,381	€ 1,657		€ 3,038
Jefe de Proyecto Implementación	29.03 hrs		Wed 6/26/19	Sat 7/6/19								€ 478	€ 574		€ 1,052
Test Implementación	88 hrs		Wed 6/26/19	Sat 7/6/19								€ 754	€ 905		€ 1,658
Soporte Implementación	23.77 hrs		Wed 6/26/19	Sat 7/6/19								€ 149	€ 178		€ 327

Task Name	Work	Duration	Start	Finish	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Total
Despliegue Centro 2	10.63 hrs	1 day	Sun 7/7/19	Sun 7/7/19									€ 246		€ 246
Jefe de Proyecto Implementación	2.63 hrs		Sun 7/7/19	Sun 7/7/19									€ 96		€ 96
Test Implementación	8 hrs		Sun 7/7/19	Sun 7/7/19									€ 151		€ 151
Seguimiento Centro 2	123.12 hrs	9 days	Mon 7/8/19	Tue 7/16/19									€ 2,545		€ 2,545
Jefe de Proyecto Implementación	23.77 hrs		Mon 7/8/19	Tue 7/16/19									€ 861		€ 861
Test Implementación	72 hrs		Mon 7/8/19	Tue 7/16/19									€ 1,357		€ 1,357
Soporte Implementación	27.37 hrs		Mon 7/8/19	Tue 7/16/19									€ 327		€ 327
Despliegue Centro 3	10.63 hrs	1 day	Wed 7/17/19	Wed 7/17/19									€ 246		€ 246
Jefe de Proyecto Implementación	2.63 hrs		Wed 7/17/19	Wed 7/17/19									€ 96		€ 96
Test Implementación	8 hrs		Wed 7/17/19	Wed 7/17/19									€ 151		€ 151
Seguimiento Centro 3	123.12 hrs	9 days	Thu 7/18/19	Fri 7/26/19									€ 2,545		€ 2,545
Jefe de Proyecto Implementación	23.77 hrs		Thu 7/18/19	Fri 7/26/19									€ 861		€ 861
Test Implementación	72 hrs		Thu 7/18/19	Fri 7/26/19									€ 1,357		€ 1,357
Soporte Implementación	27.37 hrs		Thu 7/18/19	Fri 7/26/19									€ 327		€ 327
Despliegue Centro 4	10.63 hrs	1 day	Sat 7/27/19	Sat 7/27/19									€ 246		€ 246
Jefe de Proyecto Implementación	2.63 hrs		Sat 7/27/19	Sat 7/27/19									€ 96		€ 96
Test Implementación	8 hrs		Sat 7/27/19	Sat 7/27/19									€ 151		€ 151
Seguimiento Centro 4	123.12 hrs	9 days	Sun 7/28/19	Mon 8/5/19									€ 1,131	€ 1,414	€ 2,545
Jefe de Proyecto Implementación	23.77 hrs		Sun 7/28/19	Mon 8/5/19									€ 383	€ 478	€ 861
Test Implementación	72 hrs		Sun 7/28/19	Mon 8/5/19									€ 603	€ 754	€ 1,357
Soporte Implementación	27.37 hrs		Sun 7/28/19	Mon 8/5/19									€ 145	€ 182	€ 327
Despliegue Centro 5	10.63 hrs	1 day	Tue 8/6/19	Tue 8/6/19										€ 246	€ 246

Task Name	Work	Duration	Start	Finish	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Total
<i>Jefe de Proyecto Implementación</i>	2.63 hrs		Tue 8/6/19	Tue 8/6/19										€ 96	€ 96
<i>Test Implementación</i>	8 hrs		Tue 8/6/19	Tue 8/6/19										€ 151	€ 151
Seguimiento Centro 5	195.12 hrs	9 days	Wed 8/7/19	Thu 8/15/19										€ 4,372	€ 4,372
<i>Jefe de Proyecto Implementación</i>	23.77 hrs		Wed 8/7/19	Thu 8/15/19										€ 861	€ 861
<i>Analista Implementación</i>	72 hrs		Wed 8/7/19	Thu 8/15/19										€ 1,827	€ 1,827
<i>Test Implementación</i>	72 hrs		Wed 8/7/19	Thu 8/15/19										€ 1,357	€ 1,357
<i>Soporte Implementación</i>	27.37 hrs		Wed 8/7/19	Thu 8/15/19										€ 327	€ 327

14 Conclusiones

Tras analizar las entrevistas, los diferentes orígenes de datos y el objetivo principal trasladado por la compañía Venado, se concluye que se necesita un optimizador de rutas para los cobradores a fin de ahorrar costes derivados de los múltiples desplazamientos. Asimismo, es muy importante definir indicadores de productividad medibles para conocer diariamente si las acciones aplicadas son las más adecuadas o, por el contrario, es necesario modificar algún proceso.

Se propone una solución que mediante procesos basados en algoritmos de machine learning se elaboren las rutas óptimas para los cobradores.

Una vez analizado el problema y contrastadas las hipótesis con las distintas oficinas del cliente podemos decir que:

- Existe un problema, y
- **OPTIMIZA** ha presentado una propuesta de valor desarrollando una solución a medida.

El proyecto se divide en cuatro fases:

- Extracción: obtención de los diferentes orígenes de datos.
- Dataset: Elaboración de la estructura del dataset definitivo para aplicar algoritmos de machine learning.
- Machine Learning: Aplicación de algoritmos de decisión utilizando machine learning para su ejecución.
- Industrialización – App: Encapsular la solución para que sea utilizada en cualquier parte del proceso: planificación inicial, replanificación, modificación manual.

Características de la solución:

- A implementar en una oficina piloto (La Paz).
- Una vez puesta a prueba el periodo de implementación global es de cuatro meses y medio.
- Solución tecnológica basada en los datos generados por el cliente.
- Solución automatizada.
- Los cobradores reciben sus rutas en un dispositivo móvil.
- Genera Cuadros de Mando para posterior análisis.

Como objetivo final se pretende que el sistema reduzca el fallo en las cobranzas en un 50% del valor actual (índice total de no recobro 15%), atacando las casuísticas de cliente no encontrado y cliente que no paga, por medio del estudio pormenorizado de sus patrones de conducta.

El beneficio presentado por la implementación del proyecto es positivo para Industrias Venado en tema económico, convirtiendo a la empresa en pionera en implementación de Machine Learning para planificar y realizar las cobranzas. Área muy importante de la empresa ya que más del 86% de la facturación es pagado en efectivo.

Índice de ilustraciones

Ilustración 1. Localización. Fuentes: Wikipedia, Grupo Venado	5
Ilustración 2. Volumen actividad Venado. Fuente Informe Anual Grupo Venado	6
Ilustración 3. Bolivia. Fuente: Maquetas Santa Cruz	7
Ilustración 4. Proceso. Fuente propia	10
Ilustración 5. Ubicación clientes.....	18
Ilustración 6. Resultado de las visitas de cobradores	19
Ilustración 7. Cobranzas fallidas. Fuente: Venado	19
Ilustración 8. Estructura de tablas.....	20
Ilustración 9. Ruta	22
Ilustración 10. Fases del proyecto.....	24
Ilustración 11. Modelo de Negocio proyecto optimización de rutas de cobranzas. Industrias Venado	30
Ilustración 12. Plan de acción.....	32
Ilustración 13. Histórico de revisitas	34
Ilustración 14. Evolución del índice de repetición de visitas	35
Ilustración 15. Histórico de cobranzas	35
Ilustración 16. Evolución del flujo de caja.....	36
Ilustración 17. % de Éxito en las cobranzas.....	36
Ilustración 18. Objetivo de mejora en el rendimiento de las cobranzas	38
Ilustración 19. Plan de acción.....	38
Ilustración 20. Modelo lógico.....	39
Ilustración 21. Solución tecnológica.....	40
Ilustración 22. Detalle de la solución	41
Ilustración 23. Detalle SQL. Variables	42
Ilustración 24. Detalle SQL. Creación de tabla temporal	42
Ilustración 25. Evaluación del modelo de clasificador Random-forest.....	43
Ilustración 26. Dashboard	45
Ilustración 27. Planificación de rutas	46
Ilustración 28. Organización por distribuidores.....	46
Ilustración 29. Diagrama de Bloques de la aplicación.....	47
Ilustración 30. Organización.....	48
Ilustración 31. Planificación a alto nivel.....	50
Ilustración 32. Planificación. Fuente: Elaboración Propia.....	51
Ilustración 33. Evolución de flujos de caja	57

ANEXO

Anexo I: Código de generación del dataset del modelo.

```

/*****
/* dataset de entrenamiento del modelo */
*****/
drop table tmp_venado_deudas;
drop table tmp_venado_visitas_ratio;
drop table tmp_venado_visitas_def;
drop table tmp_venado_visitas_4;
drop table tmp_venado_visitas_3;
drop table tmp_venado_visitas;
drop table tmp_venado_payment;

SELECT visit.id as visit_id,
visit.visit_date,
to_char(visit.visit_date,'d') as dayWeekVisita,
to_char(visit.visit_date,'w') as weekMonthVisita,
to_char(visit.visit_date,'wy') as weekYearVisita,
visit.employee_id,
visit.customer_id,
visit.result_id,
visit.latitude,
visit.longitude,
visit.spent_time,
visit.registration_date,
visit.operation_type,
cobradores.distributor_id,
customer.business_classification,
customer.customer_type,
customer.volume_category,
customer.regime_type,
customer.province_id,
extract(YEAR FROM age(current_date::DATE ,customer.registration_date::DATE))*12 +
extract(MONTH FROM age (current_date::DATE, customer.registration_date::DATE)) as
meses_customer,
customer_credit.line_of_credit_id,
customer_credit.credit_amount,
line_of_credit.name
into tmp_venado_visitas
from visit inner join
    venado_cobradores as cobradores on visit.employee_id = cobradores.employee_id inner
join

```

```

customer on visit.customer_id = customer.id left outer join
customer_credit on visit.customer_id = customer_credit.customer_id and
customer_credit.status = 0 left outer join
line_of_credit on customer_credit.line_of_credit_id = line_of_credit.id
where visit.operation_type in (1, 4) and
visit.origin = 1 and
customer.is_prodiverse = false and
visit.visit_date::date between '2017-01-01' and '2019-02-28';

select a.id,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 30) and
a.operation_type = 1 then 1 else 0 end) as total_visit_ko_30d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 30) and
a.operation_type = 4 then 1 else 0 end) as total_visit_OK_30d,
sum(case when a.operation_type = 1 then 1 else 0 end) as total_visit_ko,
sum(case when a.operation_type = 4 then 1 else 0 end) as total_visit_OK,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 5) and
a.operation_type = 1 then 1 else 0 end) as total_visit_ko_5d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 5) and
a.operation_type = 4 then 1 else 0 end) as total_visit_OK_5d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 10) and
a.operation_type = 1 then 1 else 0 end) as total_visit_ko_10d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 10) and
a.operation_type = 4 then 1 else 0 end) as total_visit_OK_10d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 20) and
a.operation_type = 1 then 1 else 0 end) as total_visit_ko_20d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 20) and
a.operation_type = 4 then 1 else 0 end) as total_visit_OK_20d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 60) and
a.operation_type = 1 then 1 else 0 end) as total_visit_ko_60d,
sum(case when (extract(day from a.visit_date - b.visit_date) between 0 and 60) and
a.operation_type = 4 then 1 else 0 end) as total_visit_OK_60d
into tmp_venado_visitas_ratio
from visit a,
tmp_venado_visitas b
where a.operation_type IN (1,4)
AND a.customer_id = b.customer_id
and date(a.visit_date) > date(b.visit_date)
group by a.id;

/* se incluyen las ok y ko's de los últimos 30 días */
select a.*,
coalesce(b.total_visit_ko_30d,0) as total_visit_ko_30d,
coalesce(b.total_visit_ok_30d,0) as total_visit_ok_30d,
coalesce(b.total_visit_ko,0) as total_visit_ko,
coalesce(b.total_visit_ok,0) as total_visit_ok,
coalesce(b.total_visit_ko_5d,0) as total_visit_ko_5d,

```

```

coalesce(b.total_visit_ok_5d,0) as total_visit_ok_5d,
coalesce(b.total_visit_ko_10d,0) as total_visit_ko_10d,
coalesce(b.total_visit_ok_10d,0) as total_visit_ok_10d,
coalesce(b.total_visit_ko_20d,0) as total_visit_ko_20d,
coalesce(b.total_visit_ok_20d,0) as total_visit_ok_20d
into tmp_venado_visitas_3
from tmp_venado_visitas a left outer join
  tmp_venado_visitas_ratio b on A.visit_id = b.id;

/* cuanto se recolecta por visita */
select visit_id, sum(total_payment) as total_pagado
into tmp_venado_payment
from collection
group by visit_id;

select a.*,
  coalesce(b.total_pagado,0) as total_pagado -- No es valida porque correla al 100% con
recobrado.
into tmp_venado_visitas_4
from tmp_venado_visitas_3 a left outer join
  tmp_venado_payment b on A.visit_id = b.visit_id;

--
select a.visit_id, a.customer_id,a.visit_date,
  sum(b.amount) as deuda_acumulada,
  min(b.due_date) as fecha_deuda,
  extract(day from (a.visit_date)-min(b.due_date)) as dias_deuda
into tmp_venado_deudas
from tmp_venado_visitas_4 a inner join
  debt b on a.customer_id = b.customer_id and a.visit_date > b.due_date
  and (b.payment_date is null or b.payment_date > a.visit_date)
group by a.visit_id, a.customer_id,a.visit_date;

select a.*,
  coalesce(b.deuda_acumulada,0) as deuda_acum,
  coalesce(b.dias_deuda,0) as dias_deuda
into tmp_venado_visitas_def
from tmp_venado_visitas_4 a left outer join
  tmp_venado_deudas b on a.visit_id = b.visit_id;

select * from tmp_venado_visitas_def;

```

Anexo II: Código del modelo supervisado

Modelo de predicción de cobranzas para Venado-Testing

El objetivo de este notebook es crear un modelo predictivo que nos permita ordenar los clientes de Venando con deuda pendiente por probabilidad de cobro basándonos en la información histórica de la base de datos de la empresa relativa al proceso de cobranzas. Aplicaremos un modelo predictivo basado en el algoritmo de clasificación Random Forest para estimar la probabilidad de cobro y un modelo de regresión para estimar la mejor hora del día para planificar la visita del cobrador.

Cargamos las librerías que se usarán en el modelo

In []:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import SGDClassifier
from sklearn.preprocessing import StandardScaler
```

ETL y Análisis descriptivo

1.- Cargamos los datos y convertimos el formato de las columnas:

- Fechas
- Coordenadas
- ..

2.- Convertimos employee_id y customer_id en categórica

3.- Suprimimos columnas que no aportan nada y además contienen nulos

4.- Añadimos columna nueva extrayendo el día del mes y la hora en la que se han realizado las visitas

In []:

```
visit = pd.read_csv('./visitas ampliado/visitas_ampliado_def_20190323.csv',',',
,
                    parse_dates=['visit_date','registration_date'], dtype=
{'employee_id': object, 'customer_id': object},
                    low_memory=False)

visit.head()
```

In []:

```
visit.info()
```

In []:

```
visit_etl=visit.drop(columns=['visit_id','latitude','longitude','registration_
date','result_id','name'])
visit_etl.head()
```

In []:

```
hora=pd.DataFrame({"hora": visit_etl['visit_date'].dt.hour}) #Extraemos la hor
a de la fecha de las visitas

dia=pd.DataFrame({"dia": visit_etl['visit_date'].dt.day}) #Extraemos el dia de
l mes de la fecha de las visitas

visit_proc=pd.concat([visit_etl, hora, dia], axis=1) #Creamos un nuevo datafra
me añadiendo los campos día y hora

visit_proc=visit_proc.fillna(0) #Rellenamos los valores nulos con 0

visit_proc.head()
```

In []:

```
visit_proc.info()
```

A continuación procesaremos las variables obtenidas:

- Linealizamos las variables categóricas que consideramos que tendrán influencia en el comportamiento de los clientes
- Eliminamos variables que no aportan nada
- Normalizamos las variables numéricas para igualar el peso de las variables

In []:

```
# Encontramos las variables categóricas para luego linealizarlas
obj_visit_proc = visit_proc.select_dtypes(include=['object']).copy()
obj_visit_proc.head()
```

Linealizamos (solo regime_type, ya que las otras no aportan nada al estudio)

In []:

```
visit_proc_encoding = pd.get_dummies(obj_visit_proc, columns=['regime_type'])
#Linealizamos la variable regime_type
visit_proc_encoding.head()
```

In []:

```
visit_proc_encoding=visit_proc_encoding.drop(columns=['employee_id','customer_id']) #Eliminamos employee_id y customer_id del dataset
visit_proc_encoding.info()
```

In []:

```
visit_proc_num = visit_proc.select_dtypes(include=['int64','float64']).copy()
#Seleccionamos variables numéricas
visit_proc_num.head()
```

In []:

```
visit_proc_num=pd.concat([visit_proc_num, visit_proc_encoding], axis=1) # unimos de nuevo todas las variables en un único dataset
visit_proc_num.head()
```

In []:

```
visit_proc_num.info()
```

Vamos a normalizar las variables numéricas, en primer lugar localizamos aquellas que toman valores mayores que 1 en módulo:

In []:

```
#Seleccionamos las variables cuyos valores son inferiores a -1 o mayores que 1
tt = visit_proc_num.describe().transpose()
tt[(tt['max']>1) | (tt['min']< -1)]
```

In []:

```
columns_to_norm = ['dayweekvisita','weekmonthvisita','weekyearvisita','spent_time','distributor_id','business_classification','volume_category','province_id','meses_customer','line_of_credit_id','credit_amount','total_visit_ko_30d','total_visit_ok_30d','total_visit_ko','total_visit_ok','total_visit_ko_5d','total_visit_ok_5d']
```

```
l_visit_ok_5d', 'total_visit_ko_10d', 'total_visit_ok_10d', 'total_visit_ko_20d',
'total_visit_ok_20d', 'total_pagado', 'deuda_acum', 'dias_deuda', 'hora', 'dia']
```

In []:

```
# Al normalizar las variables, los resultados que obtengamos al aplicar Random
Forest estarán también normalizados entre 0 y 1, así que calculamos el rango e
n el que está la variable hora para 'desnormalizar' al final

hora_ini=min(visit_proc['hora'])
hora_fin=max(visit_proc['hora'])
rango_hora=hora_fin-hora_ini

rango_hora # vamos a normalizar variables, por tanto al predecir la hora obten
dremos un valor entre 0 y 1 que 'desnormalizaremos' con el varlor del rango de
horas del dataset original
```

In []:

```
# Histograma para la variable hora: vemos la distribución de las visitas de co
branzas a cliente en cada banda horaria

plt.hist(visit_proc.hora, color='orange', label='Visitas')
plt.legend(loc='upper right')
plt.xlabel('Hora')
plt.ylabel('Núm. de Visitas')
plt.title('Distribución horaria de visitas')
```

In []:

```
plt.hist(visit_proc.hora[(visit_proc['operation_type']==4)], color='green', la
bel='Visita OK')
plt.hist(visit_proc.hora[(visit_proc['operation_type']==1)], color='red', labe
l='Visita KO')
plt.legend(loc='upper right')
plt.xlabel('Hora')
plt.ylabel('Núm. de Visitas')
plt.title('Resultado visitas por hora')
plt.show()
```

Y normalizamos:

In []:

```
from sklearn import preprocessing
min_max_scaler = preprocessing.MinMaxScaler()
```

```
visit_proc_num[columns_to_norm]=min_max_scaler.fit_transform(visit_proc_num[columns_to_norm])
```

Comprobamos que no quedan variables sin normalizar:

In []:

```
#Comprobamos
tt = visit_proc_num.describe().transpose()
tt[(tt['max']>1) | (tt['min']< -1)]
```

In []:

```
corr_pca = visit_proc_num.corr() # Vamos a analizar el nivel de correlación de las variables de nuestro dataset
plt.figure(figsize=(12, 10))

sns.heatmap(corr_pca,
            cmap='viridis', vmax=1.0, vmin=-1.0, linewidths=0.1,
            annot=True, annot_kws={"size": 8}, square=True);
```

In []:

```
visit_proc_num['operation_type'] = visit_proc_num['operation_type'].map({4:0, 1:1}) # Normalizamos el resultado de la visita: Operation_type=0 -> Visita OK / Operation_type=1 -> Visita KO
```

In []:

```
visit_proc_num.head()
```

In []:

```
visit_proc_num.info()
```

[Probabilidad de realizar el cobro](#)

Para estimar la probabilidad de cobro, vamos a aplicar un algoritmo de clasificación basado en Random Forest, para ello dividimos el conjunto de datos del histórico de visitas, en subconjunto de entrenamiento y test, lo que nos va a permitir evaluar el modelo

In []:

```
# *** ENTRENAMIENTO INICIAL ***
```



```

from sklearn.model_selection import train_test_split;
data_train, data_test = train_test_split(visit_proc_num, test_size=0.35)
X_train = data_train.drop(['operation_type', 'total_pagado', 'spent_time'], axis=
1) #Eliminamos las variables que no se podrán usar en el modelo predictivo ya
que su valor no está disponible antes de realizar la visita
Y_train = data_train['operation_type']

X_test = data_test.drop(['operation_type', 'total_pagado', 'spent_time'], axis=1)
Y_test = data_test['operation_type']

```

In []:

```

# Creamos una función que nos servirá para mostrar una matriz de confusión a p
artir de datos de test reales y los estimados y evaluar el modelo

from sklearn.metrics import confusion_matrix, classification_report, auc, prec
ision_recall_curve, roc_curve

def plot_confusion_matrix(y_test, pred):

    y_test_legit = y_test.value_counts()[0]
    y_test_fraud = y_test.value_counts()[1]

    cfn_matrix = confusion_matrix(y_test, pred)
    cfn_norm_matrix = np.array([[1.0 / y_test_legit, 1.0/y_test_legit], [1.0/y_t
est_fraud, 1.0/y_test_fraud]])
    norm_cfn_matrix = cfn_matrix * cfn_norm_matrix

    fig = plt.figure(figsize=(12,5))
    ax = fig.add_subplot(1,2,1)
    sns.heatmap(cfn_matrix, cmap='coolwarm_r', linewidths=0.5, annot=True, ax=ax)
    plt.title('Confusion Matrix')
    plt.ylabel('Real Classes')
    plt.xlabel('Predicted Classes')

    ax = fig.add_subplot(1,2,2)
    sns.heatmap(norm_cfn_matrix, cmap='coolwarm_r', linewidths=0.5, annot=True, ax
=ax)

    plt.title('Normalized Confusion Matrix')
    plt.ylabel('Real Classes')
    plt.xlabel('Predicted Classes')

```

```
plt.show()

print('---Classification Report---')
print(classification_report(y_test,pred))
```

Aplicamos el clasificador Random Forest a los datos de entrenamiento y después predecimos el comportamiento de los datos de test. En el caso del entrenamiento inicial para evaluar el modelo, usamos una matriz de confusión que comparará los datos de la predicción con los del resultado de la visita que teníamos del histórico.

In []:

```
# Aplicamos el clasificador a los datos de entrenamiento
from sklearn.ensemble import RandomForestClassifier
rf =RandomForestClassifier(n_estimators=100, max_depth=4, random_state=0)
rf.fit(X_train, Y_train)
Y_pred = rf.predict(X_test)

plot_confusion_matrix(Y_test, Y_pred)
```

Evaluación de la relevancia de las variables usadas en la evaluación del modelo

In []:

```
importance = pd.DataFrame(rf.feature_importances_, index=X_train.columns,
                          columns=['importance']).sort_values('importance',asc
ending=False)
importance
```

In []:

```
plt.figure(figsize=(12, 5))
plt.bar(importance.index, importance['importance'])
plt.xticks(importance.index, size = 'small', color = 'black', rotation = 90)
plt.xlabel('Variables')
plt.ylabel('% Relevancia')
plt.title('Relevancia de las variables en el resultado de la visita')
```

Obtenemos la probabilidad de éxito / ko de las visitas del subconjunto de datos de test

In []:

```
# Aplicamos el modelo predictivo para obtener la probabilidad de obtener un val
or 0 o 1 como resultado al aplicar el algoritmo a los datos de test

Y_pred_proba = rf.predict_proba(X_test)

[Y_pred_proba]
```

In []:

```
#Convertimos el array resultado en un dataframe

prevision=pd.DataFrame(Y_pred_proba, columns=['Prob_OK','Prob_KO'])

prevision.head()
```

In []:

```
prevision.info()
```

El dataset de salida será el conjunto de datos de test añadiendo la probabilidad de OK:

In []:

```
salida=X_test.reset_index().rename(columns={'index':'idx'}).drop(columns='idx'
)

salida['prob_ok']=prevision.Prob_OK

salida.head()
```

In []:

```
salida.info()
```

Mejor hora para visitar a un cliente

Para calcular la mejor hora de visita a los clientes, vamos aplicar un modelo de regresión a los datos del histórico de visita, igual que en el caso anterior. Trabajaremos únicamente con las visitas exitosas ya que se trata de intentar encontrar la mejor hora para visitar al cliente y conseguir cobrar la deuda pendiente

In []:

```
visit_proc_num_ceros = visit_proc_num.loc[visit_proc_num['operation_type'] ==
0]
```

```
visit_proc_num_ceros.shape
```

In []:

```
# *** GENERACIÓN DEL DATASET DE ENTRENAMIENTO Y TEST PARA EVALUACIÓN DEL MODELO, CON EL MISMO SPLIT DEL CASO ANTERIOR ***
from sklearn.model_selection import train_test_split;
data_train_h, data_test_h = train_test_split(visit_proc_num_ceros, test_size=0.35)
X_train_h = data_train.drop(['operation_type', 'total_pagado', 'spent_time', 'hora'], axis=1)
Y_train_h = data_train['hora']

X_test_h = data_test.drop(['operation_type', 'total_pagado', 'spent_time', 'hora'], axis=1)
Y_test_h = data_test['hora']
```

In []:

```
X_test.head()
```

In []:

```
X_test.info()
```

Aplicamos un modelo de regresión para entrenar el modelo y evaluar los resultados de la predicción en el cálculo de la mejor hora para realizar la visita a los clientes

In []:

```
from sklearn.ensemble import RandomForestRegressor
rf_h = RandomForestRegressor(n_estimators=100, max_depth=4, random_state=0)
rf_h.fit(X_train_h, Y_train_h)
Y_pred_h = rf_h.predict(X_test_h)
```

In []:

```
Y_pred_h_df = pd.DataFrame(Y_pred_h, columns=['hora_prev'])
Y_pred_h_df.head()
```

In []:

```
# Evaluamos el modelo calculando el error en la predicción mediante medios cuadrados comparando los resultados reales de la visita con la predicción
```

```

from sklearn import metrics
print('MSE '+ str(metrics.mean_squared_error(Y_test_h,Y_pred_h)))
print('MAE '+ str(metrics.mean_absolute_error(Y_test_h,Y_pred_h)))
print('RMSE '+ str(np.sqrt(metrics.mean_squared_error(Y_test_h,Y_pred_h))))

```

In []:

```

importance_h = pd.DataFrame(rf_h.feature_importances_, index=X_train_h.columns
,
                           columns=['importance']).sort_values('importance',a
scending=False)
importance_h

```

In []:

```

plt.figure(figsize=(12, 5))
plt.bar(importance_h.index, importance_h['importance'])
plt.xticks(importance_h.index, size = 'small', color = 'black', rotation = 90)
plt.xlabel('Variables')
plt.ylabel('% Relevancia')
plt.title('Relevancia de las variables en el resultado de la visita')

```

In []:

```

#salida=X_test.reset_index().rename(columns={'index':'idx'}).drop(columns='idx
')
salida['horario']=(23*Y_pred_h_df.hora_prev).round(0) #Desnormalizamos la vari
able hora_prev para añadirla al dataframe
salida.head()

```

In []:

```
salida.info()
```

In []:

```

plt.hist(Y_test_h*rango_hora, color='lightgreen', label='Resultado real')
plt.hist(salida['horario'], label='Predicción')
plt.legend(loc='upper right')
plt.xlabel('Hora de visita prevista')
plt.ylabel('Número de visitas')
plt.title('Horario de visitas OK')

```

```
plt.show()
```

In []:

In []:

Planificación de rutas

Entrenamiento continuo, será necesario actualizar la query que genera el CSV de histórico de visitas. Los datos de test en este caso son los de las visitas a planificar que se usarán para generar las rutas:

In []:

```
from sqlalchemy import create_engine
import psycopg2
import io
```

In []:

```
engine = create_engine('postgresql+psycopg2://venadoadmin:XtremeSGV16@sgv-testing.cspubeajsa0i.us-east-1.rds.amazonaws.com/venado-testing')
```

In []:

```
conn = engine.raw_connection()
cur = conn.cursor()
#cur.execute('select * from tmp_venado_propuesta_visitas_3')
```

In []:

```
# *** ENTRENAMIENTO CONTINUO ***
#X_train = visit_proc_num.drop(['operation_type', 'total_pagado', 'spent_time'],
axis=1)
#Y_train = visit_proc_num['operation_type']

# *** DATOS PARA PLANIFICAR VISITAS POR PROBABILIDAD DE ÉXITO ***
#planificacion = pd.read_csv('./visitas ampliado/visitas_propuestas.csv', ',', low_memory=False, parse_dates=['schedule_date'])
```

```
# leer de la tabla con las visitas previstas.
planificacion = pd.read_sql('SELECT * FROM tmp_venado_propuesta_visitas_3', en
gine, parse_dates=['schedule_date'])
customer=pd.DataFrame(planificacion.customer_id)
f_deuda=pd.DataFrame(planificacion.schedule_date)
planificacion=planificacion.drop(['customer_id','latitude','longitudo','busine
ss_neighborhood'], axis=1)
dia=pd.DataFrame({"dia": planificacion['schedule_date'].dt.day})
hora=pd.DataFrame({"hora": planificacion['schedule_date'].dt.hour})+12 #Usamos
como hora por defecto las 12
planificacion=pd.concat([planificacion, dia, hora], axis=1)
```

In []:

```
planificacion.head()
```

In []:

```
planificacion.info()
```

In []:

Después de cargar los datos de los clientes por fecha de vencimiento de la deuda: es decir, las rutas se calcularán para los clientes cuya deuda venza el día en el que se planifica la visita. Como se hizo en la evaluación del modelo, vamos a linealizar las variables que nos resultan más significativas

In []:

```
# Encontramos las variables categóricas para luego linealizarlas
plan_categ = planificacion.select_dtypes(include=['object']).copy()
plan_categ.head()
```

In []:

```
plan_categ_encoding = pd.get_dummies(plan_categ, columns=['regime_type'])
plan_categ_encoding.head()
```

In []:

```
plan_categ_encoding=plan_categ_encoding.drop(columns=['name'])
```

```
plan_categ_encoding['regime_type_0']=0 #En el fichero generado no hay regime_t
ype_0, ya veremos qué hacemos con esto
plan_categ_encoding.info()
```

In []:

```
plan_cat_num = planificacion.select_dtypes(include=['int64','float64']).copy()
#Seleccionamos variables numéricas
plan_cat_num.head()
```

In []:

```
plan_cat_num.info()
```

In []:

```
plan_cat_num=pd.concat([plan_cat_num, plan_categ_encoding], axis=1)
plan_cat_num.head()
```

In []:

```
plan_cat_num.info()
```

In []:

```
X_test = plan_cat_num.fillna(0) # Rellenamos los nulos con 0
```

In []:

```
#Seleccionamos las variables cuyos valores son inferiores a -1 y mayores que 1
tt = X_test.describe().transpose()
tt[(tt['max']>1) | (tt['min']< -1)]
```

In []:

```
columns_to_norm = ['dayweekvisita','weekmonthvisita','weekyearvisita','distrib
utor_id','business_classification','volume_category','province_id','meses_cust
omer','line_of_credit_id','credit_amount','deuda_acum','dias_deuda','total_vis
it_ok_30d','total_visit_ko','total_visit_ok','total_visit_ok_20d','total_visit
_ko_30d','total_visit_ko_5d','total_visit_ok_5d','total_visit_ko_10d','total_v
isit_ok_10d','total_visit_ko_20d','dia','hora']
```

In []:

```
from sklearn import preprocessing
min_max_scaler = preprocessing.MinMaxScaler()
```



```
X_test[columns_to_norm]=min_max_scaler.fit_transform(X_test[columns_to_norm])
```

In []:

```
tt = X_test.describe().transpose()
tt[(tt['max']>1) | (tt['min']< -1)]
```

In []:

```
X_test.head()
```

In []:

```
X_test.info()
```

A continuación aplicamos el modelo a los datos de entrenamiento y prueba para las visitas OK

In []:

```
from sklearn.ensemble import RandomForestClassifier
rf =RandomForestClassifier(n_estimators=100, max_depth=4, random_state=0)
rf.fit(X_train, Y_train)
Y_pred = rf.predict(X_test)
```

Y evaluamos la relevancia de las variables

In []:

```
importance = pd.DataFrame(rf.feature_importances_, index=X_train.columns,
                          columns=['importance']).sort_values('importance',asc
ending=False)
importance
```

In []:

```
plt.figure(figsize=(12, 5))
plt.bar(importance.index, importance['importance'])
plt.xticks(importance.index, size = 'small', color = 'black', rotation = 90)
plt.xlabel('Variables')
plt.ylabel('% relevancia')
plt.title('Relevancia de variables en el modelo de regresión')
```

```
plt.show()
```

In []:

```
Y_pred_proba = rf.predict_proba(X_test)
[Y_pred_proba]
```

In []:

```
prevision=pd.DataFrame(Y_pred_proba, columns=['Prob_OK','Prob_KO'])

prevision.head()
```

In []:

```
prevision.info()
```

In []:

```
salida=X_test.reset_index().rename(columns={'index':'idx'}).drop(columns='idx'
)
salida['prob_ok']=prevision.Prob_OK
salida.head()
```

In []:

```
salida.info()
```

Previsión de mejor hora de cobro

In []:

```
# *** ENTRENAMIENTO CONTINUO ***
#X_train_h = visit_proc_num.drop(['operation_type','total_pagado','spent_time'
,'hora'],axis=1)
#Y_train_h = visit_proc_num['hora']

# *** DATOS PARA CALCULAR LA MEJOR HORA PARA PLANIFICAR VISITAS ***
X_test_h=X_test.drop(['hora'], axis=1)
```

In []:

```
X_test_h.head()
```

In []:

```
X_test_h.info()
```

Entrenamos el modelo y añadimos el resultado al dataset

In []:

```
from sklearn.ensemble import RandomForestRegressor
rf_h =RandomForestRegressor(n_estimators=100, max_depth=4, random_state=0)
rf_h.fit(X_train_h, Y_train_h)
Y_pred_h = rf_h.predict(X_test_h)
```

In []:

```
Y_pred_h_df = pd.DataFrame(Y_pred_h, columns=['hora_prev'])*23
Y_pred_h_df.head()
```

In []:

```
importance_h = pd.DataFrame(rf_h.feature_importances_, index=X_train_h.columns
,
                           columns=['importance']).sort_values('importance',a
scending=False)
importance_h
```

In []:

```
plt.figure(figsize=(12, 5))
plt.bar(importance_h.index, importance_h['importance'])
plt.xticks(importance_h.index, size = 'small', color = 'black', rotation = 90)
plt.xlabel('Variable')
plt.ylabel('Relevancia')
plt.title('Relevancia de las variables en la estimación de la hora de visita')
plt.show()
```

In []:

```
# Añadimos el horario estimado al dataset de salida que habíamos obtenido al c
alcular el porcentaje de éxito en las visitas
salida['horario']=Y_pred_h_df.hora_prev
salida.head()
```

In []:

```
salida.info()
```

Añadimos a nuestro dataset los campos que habíamos eliminado anteriormente: customer y fecha de vencimiento de la deuda

In []:

```
salida=pd.concat([customer,f_deuda, salida.round()], axis=1)
salida.head()
```

In []:

```
salida.info()
```

En este punto tenemos ya el dataset original de clientes por fecha de vencimiento de la deuda con el % de éxito en la satélite. Para decidir qué clientes se incluirán en la ruta por tener una mayor probabilidad de cobro. Establecemos un umbral, de forma que las visitas que tengan un mayor porcentaje de éxito en las visitas. De esta forma, cuando se documenten los resultados de estas visitas a la base de datos de Venado, volverán a estar disponibles

In []:

```
umbral=0.87
```

In []:

```
salida_rutas = salida[(salida['prob_ok']>umbral)]
salida_rutas.head()
```

In []:

```
salida_rutas.info()
```

In []:

```
salida_rutas.to_csv('./out/salida_rutas.csv') #Salvamos los datos a un fichero CSV, la idea es que se vuelque en la tabla de visitas del ERP
```

In []:

```
plt.hist(salida_rutas['horario'])
```

```
plt.xlabel('Hora de visita prevista')
plt.ylabel('Número de visitas')
plt.title('Horario previsto para visitas con mayor probabilidad de éxito')
plt.show()
```

Preparamos el dataframe que se enviará a la base de datos de Venado.

In []:

```
planificacion.info()
```

In []:

```
#visit_etl['customer_id'] = pd.to_numeric(visit_etl['customer_id'])
planificacion=pd.concat([planificacion, customer], axis=1)
```

In []:

```
result_visit= pd.merge(salida_rutas,planificacion,on='customer_id',how='inner'
)
result_visit.head()

##### en pruebas: faltan latitud y longitud en el csv
```

In []:

```
from sqlalchemy import create_engine
import psycopg2
import io
```

In []:

```
conn = psycopg2.connect("host=sgv-testing.cspubeajsa0i.us-east-1.rds.amazonaws
.com dbname=venado-testing user=venoadmin password=XtremeSGV16")
```

In []:

```
engine = create_engine('postgresql+psycopg2://venoadmin:XtremeSGV16@sgv-test
ing.cspubeajsa0i.us-east-1.rds.amazonaws.com/venado-testing')
```

Creamos la tabla temporal de visitas.

In []:

```
###cambiar salida_rutas por result_visit cuando tengamos todos los campos.  
  
salida_rutas.head(0).to_sql('tmp_propuesta_visitas', engine, if_exists='replace', index=False) #truncates the table
```

Cargamos la tabla con el resultado del modelo.

In []:

```
conn = engine.raw_connection()  
cur = conn.cursor()  
output = io.StringIO()  
salida_rutas.to_csv(output, sep='\t', header=False, index=False)  
output.seek(0)  
contents = output.getvalue()  
cur.copy_from(output, 'tmp_propuesta_visitas', null='') # null values become ''  
conn.commit()
```