# **Data Science Framework Report**

# Business question

**What is the stated business question?**

¿Es posible predecir si un cliente tendrá alta probabilidad de incumplimiento de pago?

**What is the intent underlying the question (e.g., what is the context, what is the impacted segment, and what are stakeholders’ current thoughts about the underlying reasons?**

**Contexto:**

Credit One está enfrentando una problemática de credibilidad por parte de sus clientes debido a que se están otorgando créditos a personas con alto riesgo de incumplimiento de pago que Credit One no está identificando como cliente de riesgoso.

**Segmento Impactado:**

Credit One y sus clientes.

**Pensamientos de los stakeholders:**

Por parte de los dirigentes de Credit One, existe una gran preocupación respecto a esta situación, ya que, de no resolverse, se podría dar una **pérdida** significativa de clientes potenciales para la cuenta.

Por parte de los clientes afectados, de no resolverse el problema de credibilidad, se **terminará la relación** con Credit One como su proveedor de servicio de calificación crediticia.

**What business considerations (e.g., stakeholders, timeline, and cost) are likely to impact the analysis?**

Los siguientes factores podrían tener un impacto en el desarrollo del proyecto:

1. Los stakeholders solicitan una solución al problema planteado en un plazo máximo de dos semanas. Lo que podría impactar la ejecución de este, ya que es un muy corto plazo para todas las actividades a ejecutar.
2. La presión ejercida por los clientes de Credit One puede influir en la precisa toma de decisiones por parte de los stakeholders.

# Analysis plan

**What is the analysis goal?**

Diseñar, crear e implementar una solución que permita calcular el puntaje de crédito de cada cliente, haciendo uso del lenguaje de programación Python para reducir el riego de otorgar créditos a clientes con alta probabilidad de incumplimiento.

Mantener la credibilidad de los socios en el uso de Credit One como su proveedor de servicio de calificación crediticia.

**What hypotheses are to be tested?**

Los clientes con un menor credit score son mas propensos a incumplimiento de pago.

**What data is required/available to test the hypotheses?**

La información disponible para realizar el análisis es la siguiente:

Gender (1 = male; 2 = female).

Education (1 = graduate school; 2 = university; 3 = high school; 0, 4, 5, 6 = others).

Marital status (1 = married; 2 = single; 3 = divorce; 0=others).

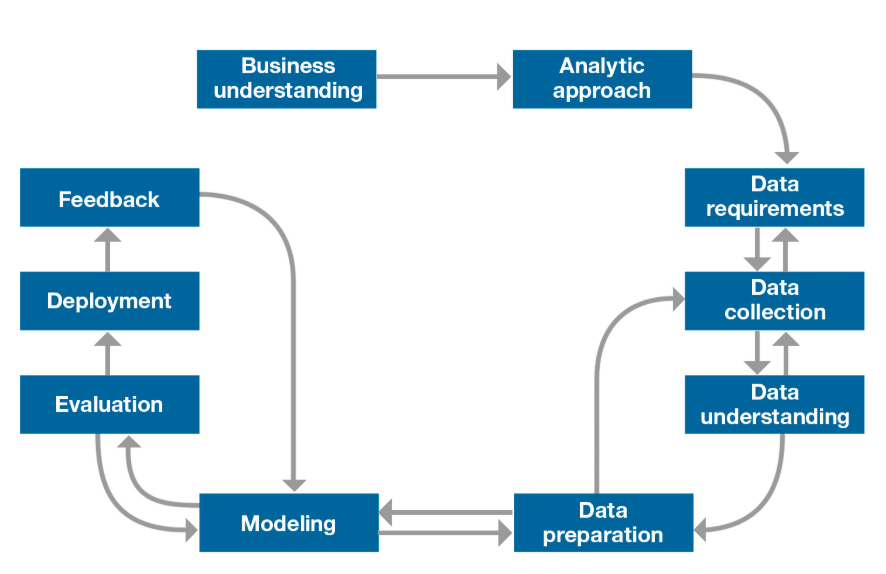
Age (year).

History of past payment. We tracked the past monthly payment records (from April to September 2005).

Client's behavior: Y=0 then not default, Y=1 then default"

**What methodology(-ies) will you employ?**

La metodología a emplear para desarrollar este proyecto es la siguiente: [**Foundational Methodology for Data Science**](http://www.slideshare.net/JohnBRollinsPhD/foundational-methodology-for-data-science). La misma, propia de IBM y consiste en las siguientes fases:



Para conocer mas detalles sobre cada una de las etapas, puede referirse al siguiente enlace:

<https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science>

**What is the project plan (timeline and milestones, risks, phasing, prioritization, …)?**

El tiempo definido por los stakeholders para el primer entregable es de 2 semanas.

Pasadas las dos semanas, se hará entrega del modelo creado y las predicciones realizadas.

Las fases por desarrollar son las siguientes, en este orden:

Las primeras fases ya han sido cubiertas por el negocio, quien ha facilitado en set de datos a usar.

*Data Understanding*: En esta fase se hará la revisión y entendimiento de los datos que servirán de fuente de datos para construir el modelo.

*Data Preparation:* Posteriormente, una vez habiendo reconocido e inspeccionado los datos, se procede a realizar conversiones y ajustes necesarios a los datos. En caso de identificar que hay atributos que no aportan valor al modelo, se eliminan eso atributos del set de datos.

*Modeling*: En esta fase, se procede a crear el modelo. Haciendo uso de los datos buscaremos la mejor combinación de atributos que nos ayuden a crear un modelo eficiente en busca de los objetivos planteados. Incluiremos las diferentes reglas ya definidas por el negocio para la clasificación de los clientes.

*Evaluation:* En esta etapa se analiza el rendimiento de los diferentes parámetros que hayamos aplicado al modelo para identificar y seleccionar el más eficiente.

*Deployment:* Una vez con el modelo completado, se procede a realizar el lanzamiento a producción, donde será evaluado con datos reales.

Feedback: Se hacen las evaluaciones por parte de los stakeholders y partes involucradas con el fin de identificar si se debe hacer ajustes al modelo.

# Data collection

**From where can the data be obtained?**

La información será facilitada por cada uno de los clientes de Credit One. Estos facilitaran la información correspondiente a la cartera de clientes de crédito que manejan, con el fin de obtener el credit score de cada uno de ellos, y usarlo en la toma de decisiones.

**How must the data be cleansed and validated?**

Se debe realizar una revisión a cada atributo con el fin de identificar si posee el tipo de dato correcto. En caso de no ser así, se debe realizar la conversión respectiva para enviar al modelo cada atributo con el tipo de dato adecuado. También se debe evaluar si todas las variables son requeridas y generan valor. De no ser así, proceder a remover las que consideremos innecesarias.

# Insights

**What patterns do you see in the data?**

Muchas de los atributos son categóricos (Gendes, Education, Sex)

Los números negativos en los atributos referentes a estatus de pagos son indicadores de que es un cliente cumplidor con sus pagos.

La mayoría de los clientes son mujeres, con un número similar de solteras y casadas.

Una cantidad reducida de clientes han terminado la universidad.

**Are each of the hypotheses proven or disproven?**

**How much confidence should stakeholders place in the results?**

El modelo será construido basado en los datos que representan el historial crediticio del cliente, por lo que el credit score proporcionado es de suma confianza y podrá ser usado en el proceso de evaluación de cada cliente al momento de otorgar un crédito.

**How do you rank your findings in terms of quantified impact on the business?**

Representa un **alto** impacto al negocio, ya que de este resultado depende la pérdida o continuidad de muchos de los clientes potenciales para Credit One.

Traducido a números podemos ver un fuerte impacto económico, así como de prestigio y la credibilidad a Credit One como proveedora de servicios.

# Recommendation

Una vez concluido el modelo y con los resultados obtenidos. Mostraremos diferentes visualizaciones a los interesados, donde se pueda comparar el porcentaje de clientes clasificados como de **alto riesgo**, así como el porcentaje de clientes considerados **excepcionales**. Esto tomando en cuenta las reglas de negocio ya establecidas en cuanto al valor del Credit Score de los clientes según Credit One.

* **Exceptional**: 800-850
* **Very Good**: 740-799
* **Good**: 670-739
* **Fair**: 580-669
* **Poor**: 300-579

Ejemplo: