# Proyecto de análisis de datos

## Analista de datos:

## Eduardo Fernández Dionicio

## Cliente/Patrocinador:

## Entidad financiera banca/fintech con cartera pyme

## Propósito:

Este proyecto tiene como objetivo investigar el riesgo de crédito en préstamos SBA 7(a) (EE. UU., 2020–2025) con el fin de mejorar la decisión de originación. Identificaremos segmentos de mayor riesgo por industria (NAICS), geografía (Estado), tamaño del préstamo y plazo, y construiremos un modelo logístico (PD) sencillo y explicable que ordene a los solicitantes por probabilidad de incumplimiento.

Con base en ese ranking (deciles), definiremos un umbral de acción (cut-off) para priorizar revisión/ajustes de límite o plazo y estimaremos el impacto esperado sobre charge-off y pérdida esperada (EL ≈ PD × LGD\_proxy × EAD). El resultado se comunicará con KPIs de cohortes/vintages, mapas/ tablas por segmento y un memo ejecutivo con recomendaciones accionables para Riesgos, Originación y Dirección.

## Alcance / Actividades principales del proyecto:

|  |  |
| --- | --- |
| Actividad | Descripción |
| Recogida de datos | Recolectar datos del dataset SBA 7(a) FOIA 2020–presente |
| Ordenar por Probabilidad de Default | Ordenar solicitudes por riesgo con un PD simple |
| Identificar deciles con mayores impagos | Analizar los deciles concentran los impagos y poner un umbral que active revisión/ajustes |
| Crear dashboard con trade-off riesgo-volumen | Presentar dashboard claro mostrando el trade-off entre riesgo y volumen |
| Entregar reporte final | Entrega de reporte final con recomendaciones para la entidad financiera. |

## Este proyecto no incluye:

* Cualquier área fuera de los límites de los Estados Unidos.
* Implementaciones de cualquier solución o recomendación.
* Datos de préstamos de más de 5 años en el proyecto (2020-2025)

## Entregas:

|  |  |
| --- | --- |
| Entregable | Descripción/ Detalles |
| Scope of Work | Una declaración clara de la tarea empresarial que has seleccionado investigar |
| Fuentes recolectadas | Una descripción de todas las fuentes de datos utilizadas |
| Documentación de limpieza o manipulación de datos | Documentación de cualquier limpieza o manipulación de datos |
| Documentación de hallazgos | Un resumen de tu análisis |
| Dashboard y hallazgos clave | Visualizaciones de apoyo y hallazgos clave |
| Documentación extra | Según lo que descubras, una lista de entregables adicionales que consideres útiles incluir para una exploración más profunda |
| Reporte final | Tus principales ideas de alto nivel basadas en tu análisis |

## Resumen del cronograma / Principales hitos:

|  |  |  |
| --- | --- | --- |
| Hito | Fecha de finalización prevista | Descripción/Detalles |
| *Revisión de datos* | *24/08/25* | *Revisión de las fuentes de datos completada* |
| *Análisis de datos* | *24/08/25* | *Análisis de datos inicial completado* |
| *Lista de deciles* | *25/08/25* | *Deciles con mayor tasa de impagos identificado* |
| *Creación de umbral* | *25/08/25* | *Umbral de revisión/ajuste establecido* |
| *Visualización de datos* | *25/08/25* | *Dashboard con trade-off y KPIs principales completado* |
| *Lista de recomendaciones* | *26/08/25* | *Lista de recomendaciones para cada decil* |
| *Reporte final* | *27/08/25* | *Reporte final detallando todo el trabajo realizado, análisis, metodologías, y descubrimientos.* |

## \*Fecha estimada de finalización:

27 de Agosto 2025