

**UNIVERSIDAD NACIONAL DE COLOMBIA
PROGRAMA CURRICULAR DE ADMINISTRACIÓN DE EMPRESAS
FACULTAD DE CIENCIAS ECONÓMICAS
SEDE BOGOTÁ**

INFORME DE PASANTÍA

Informe de trabajo de grado-Pasantía

MANUEL FELIPE RODRÍGUEZ GUZMÁN

**DIRECTOR
PEDRO JULIAN RAMÍREZ ANGULO**

BOGOTÁ D.C., AGOSTO 2025

TABLA DE CONTENIDO

1. Introducción.....	2
2. Problema	4
3. Objetivos.....	5
4. Productos obtenidos.....	6
5. Trabajos citados	12

Introducción

El siguiente informe es el resultado de la asignatura trabajo de grado en modalidad de pasantía cursado durante el segundo semestre de 2025. La modalidad de pasantía es una opción de grado a través de la cual el estudiante puede vincularse a una organización con el fin de emplear el conocimiento adquirido durante su proceso de formación para lograr una mejora clara dentro de la empresa.

Para el estudiante, la pasantía es una oportunidad para afianzar su conocimiento en un área en particular a la par que adquiere experiencia laboral en ese campo, y para la empresa, es un espacio para identificar fallas o problemas en sus procesos operativos y administrativos.

Particularmente, en el área de analítica la experiencia laboral es esencial, y si se refuerza con un portafolio de logros prácticos, se convierte en un factor diferencial en el perfil laboral de una persona. Es por eso, por lo que la modalidad de pasantía resulta fundamental como última materia del pregrado y como primera experiencia laboral en el área.

La pasantía se desarrolló en la organización Cash Block SAS, una Fintech que nace como proyecto de una casa de desarrollo de Software que usualmente se dedica al desarrollo del *core* bancario para otras financieras. Este proyecto empezó su operación en el año 2024 con el objetivo de bancarizar a la población excluida de la banca tradicional mediante soluciones tecnológicas.

Cash Block es una empresa dedicada a la colocación de crédito, y su propuesta de valor es desarrollar productos de crédito personalizados con el fin de siempre presentar una alternativa, o un cómo sí, acorde a las características del solicitante. La puerta de entrada al mercado crediticio de Cash Block es el crédito para la adquisición

de motocicletas, motivo por el cual actualmente esta es la única línea de crédito disponible.

Durante el 2024, la empresa tercerizó el servicio de colocación de crédito, y se centró en el desarrollo tecnológico para la gestión de cartera y cobranza, sin embargo, como parte del plan de expansión para el 2025 se busca desarrollar un esquema de operación totalmente autónomo. A raíz de esta decisión este año la capacidad operativa de la empresa ha crecido, a nivel de planta se han contratado pasantes para apoyar el proceso de desarrollo de producto, a nivel comercial se han cerrado alianzas con concesionarios de diversas marcas en la ciudad de Bogotá y nivel de alianzas estratégicas se han cerrado convenios con buros de crédito y con otras entidades financieras para el intercambio de datos.

Dicho lo anterior, la pasantía se desarrolló en el área de evaluación de riesgo. Dado que la empresa recién estaba desarrollando un motor de decisión propio cuando la pasantía empezó, la toma de decisiones se limitaba a estimar el riesgo en función de la capacidad de pago, sin embargo, mucha de la información del cliente, obtenida en el onboarding comercial, no se emplea de forma objetiva para la toma de decisiones.

A comienzo de año, la empresa cerró un convenio con la plataforma Carro Ya y con la marca Hero, razón por la cual se tomó la decisión de complementar al equipo con la contratación de pasantes. Dentro de las funciones del cargo, se deben generar informes semanales de la operación de colocación y gestionar las bases de datos necesarias para la generación de esos informes.

En el transcurso de la pasantía se identificó que la toma de decisiones, basada en la evaluación de riesgo, dependía de la obtención y análisis de soportes presentados por el solicitante, lo cual genera fricción y largos tiempos de espera entre onboarding y la aprobación de un crédito. Sin embargo, un modelo basado en datos demográficos y psicográficos debería poder optimizar el proceso de toma de decisiones, en algunos casos o para algunos perfiles, lo que permitiría minimizar la fricción y aumentar la tasa de conversión de aprobaciones.

La construcción de un modelo para estimar la probabilidad de impago, o para clasificar según atributos, es una tarea demasiado específica, y no es algo que se vea en concreto dentro de una materia del pregrado de Administración de empresas, no obstante, a través del pregrado si se adquieren los conocimientos y capacidades necesarios para entender la amplia bibliografía acerca del tema. Este trabajo de grado toma como referente el libro de Baesens, Roesch & Scheule (2016), *redit risk analytics: Measurement techniques, applications, and examples in SAS*. John Wiley & Sons, un libro dedicado a la construcción de modelos para la medición del riesgo de default, de

ahí se toma la metodología de trabajo. Adicionalmente, para la adopción de variables psicométricas se parte del artículo de Djeundje, Crook, Calabrese & Hamid (2021), el cual justifica la incorporación de este tipo de data en los modelos tradicionales, además se hace uso de diversos cuestionarios psicométricos, con su respectiva bibliografía, para la selección e incorporación de estas variables.

Conclusiones y recomendaciones ...

Problema

Cuando se evalúa la viabilidad de un crédito, la determinación final se toma con base en la valoración de riesgo atribuida o estima para el solicitante, y cuando hablamos de riesgo en un crédito el concepto es muy amplio, pues, abarca el riesgo de incapacidad o muerte, de desempleo, de accidente, de fraude, de suplantación y el riesgo de impago, entre muchos otros.

Para la mayor parte de los riesgos existen pólizas o seguros que cubren al deudor y al acreedor, y para el riesgo de suplantación o fraude existen validaciones biométricas y consultas en listas de control públicas que mitigan el peligro. Para el riesgo de impago o de Default sin embargo no hay una garantía definitiva que proteja al acreedor por completo, razón por la cual el problema se ha centrado en analizar tres al deudor en tres dimensiones distintas, su capacidad de pago, su voluntad de pago y las posibles garantías que mitiguen la pérdida esperada, Gitman & Zutter (2012), de acuerdo con los estándares actuales, amplían las dimensiones tradicionales agregando las condiciones económicas y el patrimonio del solicitante.

Profundizando un poco más en las dimensiones de análisis tradicionales, la capacidad se entiende como la suficiencia que tiene el deudor para cumplir con sus obligaciones de forma oportuna sin comprometer una parte significativa de su ingreso, esta dimensión se mide con los soportes presentados al momento del onboarding tales como extractos, cartas laborales, etc. La voluntad mide la disposición del deudor para realmente cumplir con la obligación, tradicionalmente son los burós de crédito

quienes se encargan de medir la voluntad mediante scores formulados a partir del comportamiento pago histórico, y finalmente las garantías son los mecanismos a través de los cuales puede el acreedor asegurar una parte de la deuda como una prenda o un pagaré.

Durante el trabajo de campo de la pasantía, hecho en concesionarios, se descubrió que la banca tradicional pondera la voluntad como el atributo de mayor peso en su proceso de decisión, un buen score en burós de crédito es suficiente para acceder a crédito en una entidad bancaria tradicional. Sin embargo, para una Fintech que tiene como propuesta de valor la inclusión financiera, es necesario que la voluntad de pago sea evaluada de forma alternativa, ya que los solicitantes que llegan son aquellos ya descartados por los grandes bancos.

Si la organización logra una media diferente, propia, de la voluntad de pago podría aprovechar la data de los solicitantes para no tener que depender únicamente de la valoración de los burós, esto generaría dos claras ventajas, el primero reducir la fricción sin aumentar significativamente la exposición al riesgo y el segundo aumentar la tasa de colocación logrando captar clientes con capacidad y voluntad que otros bancos pasan por alto. Ahí está la importancia de solicitar el problema, para los clientes estamos generando acceso al sector financiero formal, y para la organización estamos aumentando su cobertura dentro del objetivo sin que esto implique exponerse de forma excesiva al riesgo de impago.

Objetivos

Objetivo general: desarrollar un modelo que permita estimar la probabilidad de default, o bien un modelo con el cual clasificar a los solicitantes en niveles de riesgo, según sus características, con el fin de reducir la fricción sin aumentar el riesgo, a la vez que se minimizan los costos operativos de análisis y se aumenta la tasa de conversión.

Objetivos específicos:

- Depurar y ajustar la base de datos de colocación con el fin de entrenar un modelo logístico o un modelo de clasificación (XGBoost), de acuerdo con la metodología propuesta por el texto de referencia.
- Construir diferentes modelos con la base de datos resultante del objetivo anterior, para comparar las métricas precisión y desempeño.
- Realizar backtesting para cuantificar el impacto económico esperado del modelo en la cartera histórica.

- Definir metodología para implementar el modelo resultante en la toma de decisiones como medida complementaria de la voluntad de pago, no se espera que el modelo se tome como único factor decisivo, pero si pudiera tener alguna ponderación en la decisión final.
- Identificar segmentos de solicitantes donde el modelo aporte mayor valor discriminativo para optimizar decisiones automatizadas.
- Definir las variables psicométricas que puedan ayudar a estimar la voluntad de pago de forma alternativa, y desarrollar una forma de capturar variables psicométricas relevantes en el proceso de onboarding.

Productos obtenidos

1) Base de datos depurada:

De acuerdo con el libro de Baesens, Roesch & Scheule (2016), en el capítulo 4 *Data Preprocessing for Credit Risk Modeling*, la metodología de trabajo para preparar la base de datos para la construcción de modelos de riesgo es:

- **Tratamiento de datos atípicos**
- **Tratamiento de valores faltantes**
- **Estandarización (escalares)**
- **categorización de variables (categóricas)**
- **Weight of Evidence (WOE)**
- **Selección de variables**

Sin embargo, para este trabajo se tomó la decisión de reducir la metodología a: tratamiento de datos atípicos, de valores faltantes y selección de variables.

Antes de empezar con la preparación de los datos para la construcción de modelos, es necesario realizar un primer análisis exploratorio para identificar las variables, el tipo de variable y los problemas en los registros de la base.

El primer paso fue crear un diccionario de variables, el cual identifica la variable, la describe brevemente y le asigna una tipología:

VARIABLE	DESCRIPCIÓN	TIPO
number	Identificador del registro	Escalar
dtsscore	Puntaje DataCrédito del cliente principal	Escalar

dtcoreco	Puntaje DataCrédito del codeudor	Escalar
edad	Edad del cliente	Escalar
IBC	Score interno, estimado con la capacidad de pago calculada (1-10)	Escalar
labor	Categoría (empleado/independiente)	Categoría
calificacion	Clasificación del cliente según riesgo, estimado por CIFIN	Categoría
ingresos	Ingresos mensuales declarados del cliente	Escalar
pendeuda	Porcentaje de endeudamiento, estimado por DataCrédito	Escalar
co	Codeudor como garantía (1/0)	Categoría
credito	Monto total del crédito	Escalar
yeardde	Año de desembolso	Escalar
tinteres	Tasa de interés aplicada al crédito	Escalar
ncuotas	Número de cuotas pactadas	Escalar
genero	Género del cliente	Categoría
cuota	Valor de la cuota mensual	Escalar
capacidad	Capacidad de pago (cuota/ingresos)	Escalar
antiguedad_laboral	Años en el trabajo actual	Escalar
estrato	Estrato socioeconómico	Categoría
tipo_vivienda	Tipo de tenencia de vivienda	Categoría
nivel_educativo	Nivel educativo alcanzado	Categoría
creditos_activos	Número de créditos activos	Escalar
personas_hogar	Número de personas en el hogar	Escalar
eactual	Estado del crédito (0=Al día, 1=Default)	Categoría

Tabla 1

Después, en este primer análisis de la base de datos de créditos desembolsados, la cual fue entregada por el tercero que estaba realizando el proceso de colocación, se identificaron los siguientes problemas:

Categorías

- La variable “labor” tenía errores de digitación en sus registros:

```
names(table(cred$labor))
] "EMPLEADO" "EMPLEADO " "INDEPENDIENTE" "INDEPENDIETE"
```

Ilustración 1

- En variable “co” encontramos errores de digitación, además del valor “X”, el cual sabemos que implica la existencia de la garantía. También hay valores vacíos “”, los cuales se pueden eliminar mirando el valor la variable “dtcoreco”, ya que la existencia de un score asignado implica o niega la garantía.

```
> names(table(cred$co))
[1] "" "NO" "NO " "SI" "X"
```

Ilustración 2

- La variable “genero” también tiene errores de escritura

```
> names(table(cred$genero))  
[1] " HOMBRE" " MUJER"  "HOMBRE"  "HOMBRE " "MUJER"  "MUJER "
```

Ilustración 3

- La variable “estrato” tiene una categoría que corresponde a valores faltantes:

```
> names(table(cred$estrato))  
[1] "- " "1" "2" "3" "4"
```

Ilustración 4

Esto se debe a que el solicitante no presentó el soporte correspondiente, la categoría “-” se puede eliminar imputando ya que son valores nulos.

- La variable “eactual” está correctamente categorizada, sin embargo, para la construcción de modelos reduciremos las categorías a 1 (impago) y 0 (pago). Para evitar incluir eventos situacionales en el comportamiento de impago se considerará impago a partir de los 60 días.
- Las variables “calificacion”, “genero”, “tipo_vivienda”, y “nivel_educativo” no tienen categorías erróneas.

Escalares

- La variable “dtscore” no es completamente numérica

```
> class(cred$dtscore)  
[1] "character"
```

Ilustración 5

Se identificó que algunos registros tienen texto

```
"codeudor 639" "codeudor 704" "codeudor 741" "codeudor 755"  
"codeudor 771" "codeudor 779" "codeudor 784" "codeudor 794"  
"codeudor 817" "codeudor 860" "DEUDOR 0" "DEUDOR 683"  
"DEUDOR 718" "DEUDOR 722" "DEUDOR 765" "DEUDOR 771"  
"DEUDOR 0" "DEUDOR 323" "DEUDOR 503" "DEUDOR 587"  
"DEUDOR 592" "DEUDOR 628" "deudor 661" "deudor 665"  
"DEUDOR 676" "DEUDOR 682" "DEUDOR 683" "deudor 688"  
"DEUDOR 691" "DEUDOR 692" "DEUDOR 693" "deudor 695"  
"DEUDOR 695" "DEUDOR 697" "DEUDOR 698" "DEUDOR 699"
```

Ilustración 6

Por lo cual es necesario extraer la parte numérica de los registros “DEUDOR ...” y de los registros “codeudor ...”, además, El score de DataCrédito solo puede tomar valores entre 150 y 950, por ello los valores que son 0, los nulos, y los valores que tienen texto

representan un valor no válido para esta variable, tanto para deudor como para codeudor.

La existencia de un valor 0 o de un valor nulos puede tener tres causas, la primera es que la persona no tenga historial de crédito en ese caso no se genera Score, la segunda es porque el cliente tiene todas sus cuentas en Default (mora mayor a 90 días), y la tercera sería por un error al momento de registrar el valor. Para este análisis se toma por supuesto que la causa de un valor 0 o un valor nulo responde a una de las 2 primeras razones.

Ya que conocemos las causas de los valores “atípicos” en la variable sabemos que no es correcto imputar a través de un método de estimación estadístico ya que no representa la lógica real detrás de la ausencia de valor. Para facilitar la construcción de modelos que logren captar esta particularidad los valores faltantes en el score se imputaran con el valor “-999”, y crearemos una variable auxiliar, con valores 0 y 1, que identifique cuando el score fue asignado.

- Las variables “IBC”, “tinteres” y “capacidad”, que son porcentajes son del tipo cadena.
- Las demás variables escalares no tienen errores en el registro.

La depuración de la base de datos, para corregir los errores listados, se realizó en Excel ya que es una herramienta que está muy optimizada para este tipo de tarea.

Tras la depuración ahora si es posible definir correctamente las categóricas y las escalares dentro de R, de forma que la función summary, identifique todas las variables de forma correcta, según el diccionario de variables.

De la función summary, a simple vista, se identifica, en la variable “IBC” hay un dato que no está dentro del rango permitido para la variable, en la variable “ingreso” hay un valor que probablemente este sea un error de digitación y hay 41 registros nulos, en la variable “estrato” hay 37 valores nulos, la variable “credito” tiene un registro por sesenta millones cuando el monto máximo prestado por regla de negocio es de veinte millones.

Finalizada la depuración es posible empezar con el análisis de valores atípicos en las variables escalares.

Tal como se identificó en la función summary, la variable “crédito”, la variable “IBC” y la variable “ingreso” tienen al menos un registro atípico:

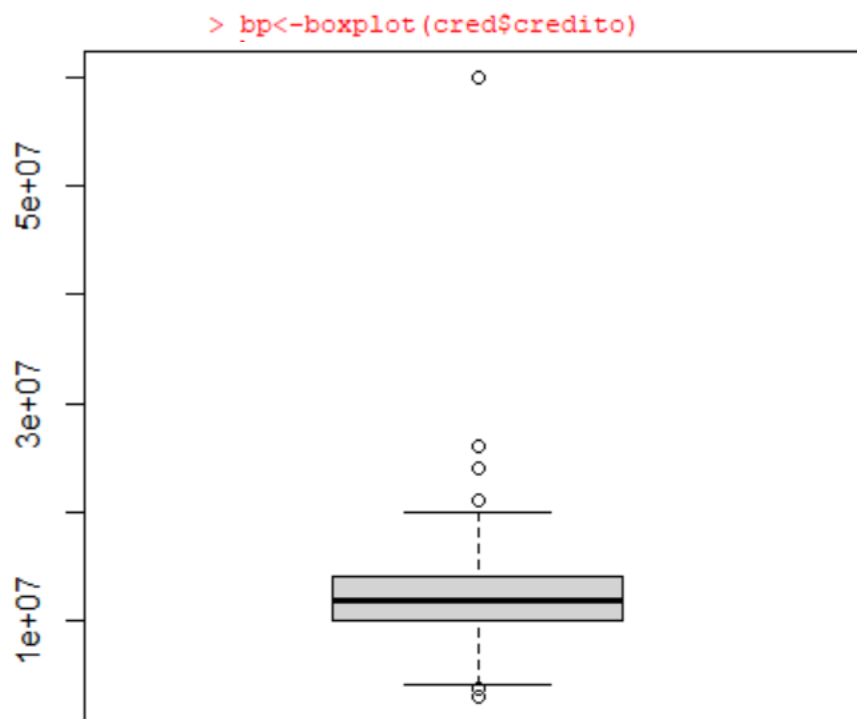


Ilustración 7

```
> bp<-boxplot(cred$ingreso)
```

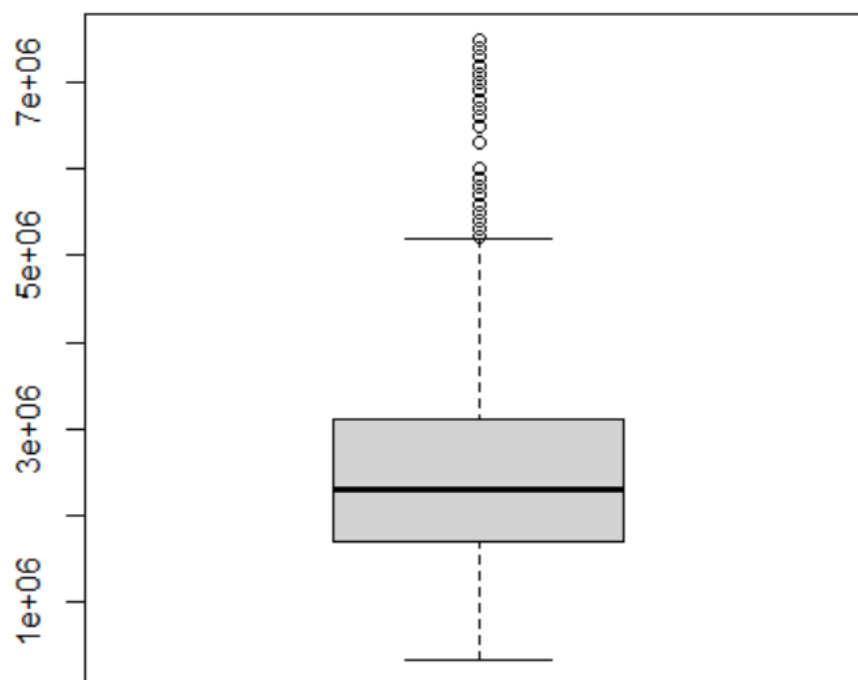


Ilustración 8

```
> bp<-boxplot(cred$IBC)
```

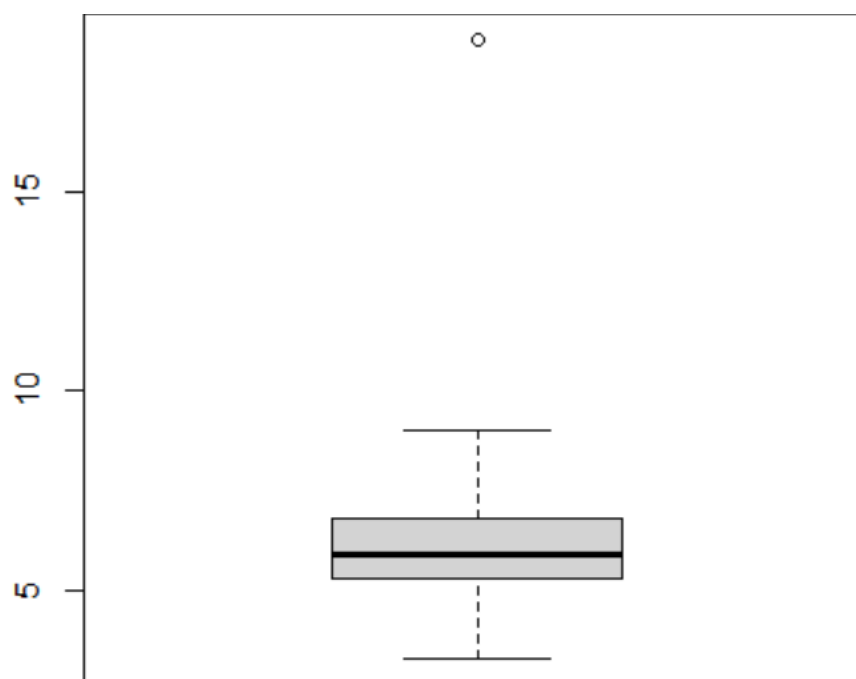


Ilustración 9

Tras la aplicación de la regla empírica se detectaron 6 registros atípicos para la variable “crédito” y otros 2 registros para la variable “IBC”, los cuales son registros diferentes, es decir los dos registros atípicos de “IBC” no están dentro de los 6 registros atípicos de “crédito”, mientras que de ingresos no se identificaron valores atípicos. Como resultado de la eliminación de valores atípicos se eliminaron 8 valores.

Trabajos citados

Gitman, L. J., & Zutter, C. J. (2012). *Principles of Managerial Finance*. 13e.

Baesens, B., Roesch, D., & Scheule, H. (2016). *Credit risk analytics: Measurement techniques, applications, and examples in SAS*. John Wiley & Sons.

Djeundje, V. B., Crook, J., Calabrese, R., & Hamid, M. (2021). Enhancing credit scoring with alternative data. *Expert Systems with Applications*, 163, 113766.

Ilustración 1	7
Ilustración 2	7
Ilustración 3	8
Ilustración 4	8
Ilustración 5	8
Ilustración 6	8
Ilustración 7	10
Ilustración 8	11
Ilustración 9	11