

# **Bitácoras Grupo #5, CA-204 (II-2024)**

**Análisis de las variables cualitativas en relación al riesgo crediticio alrededor del mundo**

Jeikel Navarro Solis, Gabriel Valverde, Erick Venegas

2024-10-25

# Tabla de contenidos

<b>Introducción</b>	<b>4</b>
<b>1 Bitacora 1</b>	<b>5</b>
1.1 Parte de planificación . . . . .	5
1.1.1 Definición de la idea . . . . .	5
1.1.2 Conceptualización de la idea . . . . .	5
1.1.3 Identificación de tensiones . . . . .	6
1.1.4 Reformulación de la idea en modo preguntas . . . . .	6
1.1.5 Argumentación de las preguntas . . . . .	6
1.1.6 Argumentación a través de datos . . . . .	9
1.2 Revisión bibliográfica . . . . .	10
1.2.1 Búsqueda de bibliografía . . . . .	10
1.2.2 Fichas de literatura . . . . .	10
1.3 Construcción de la UVE de Gowin . . . . .	15
1.3.1 Conceptos básicos . . . . .	15
1.3.2 Principios y teorías . . . . .	16
1.4 Parte de escritura . . . . .	16
<b>2 Bitácora 2</b>	<b>18</b>
2.1 Parte de Planificación . . . . .	18
2.1.1 Ordenamiento de la Literatura . . . . .	18
2.2 Enlaces de la Literatura . . . . .	19
2.3 Análisis Estadístico . . . . .	22
2.3.1 Análisis Descriptivo . . . . .	22
2.4 Propuesta Metodológica . . . . .	43
2.5 Construcción de Fichas de Resultados . . . . .	44
<b>3 Bitácora 3</b>	<b>48</b>
3.1 Parte de Planificación . . . . .	48
3.1.1 Análisis de modelación . . . . .	48
3.1.2 Construcción de fichas de resultados . . . . .	48
3.1.3 Ordenamiento de los elementos de reporte . . . . .	48
3.2 Parte de escritura . . . . .	48
3.3 Parte de reflexión . . . . .	48

<b>4</b>	<b>Anexo</b>	<b>49</b>
4.1	Anexo 1 (CHANGELOG Bitacora 1) . . . . .	49
4.1.1	Chore . . . . .	49
4.1.2	Feat . . . . .	49
4.1.3	Fix . . . . .	50
4.2	Anexo 2 (Participacion Bitacora 1) . . . . .	50
4.3	Anexo 3 (CHANGELOG Bitacora 2) . . . . .	51
4.4	Anexo 4 (Participacion Bitacora 2) . . . . .	52
<b>5</b>	<b>Apendice</b>	<b>53</b>
5.1	Correcciones Bitácora 1 . . . . .	53
5.2	Sugerencias Bitácora 1 . . . . .	53
5.3	Referencias bibliográfica . . . . .	54

# Introducción

Este estudio utiliza una base de datos de riesgos financieros en el cual se toman en cuenta variables como lo son la edad, genero, pais en el que vive y estado marital de las personas, esto con el fin de realizar un perfil descriptivo de los prestatario, asi como tambien se toman variables cuantitativas como los ingresos, activos y deuda de la persona en cuestion. De manera que la base contiene un total de 15000 casos y 20 variables. El objetivo es generar empiricamente un metodo que basado en las variables permita crear una calificación de riesgo, tal que se puedan relacionar las variables cuantitativas y cualitativas de modo tal se logre llegar a algún resultado satisfactorio a partir de estos datos. Por estas razones, la investigación se basa en un marco teorico y empirico donde los autores buscan salir de la norma y explorar mas factores que solo la parte economica de los prestatarios.

# 1 Bitacora 1

## 1.1 Parte de planificación

### 1.1.1 Definición de la idea

Cuando una persona quiere acceder a un préstamo bancario, es común que los bancos requieran información personal y profesional de la persona, por ejemplo su profesión, ingresos mensuales, edad, gastos mensuales, etc. Esto se debe a que los bancos necesitan saber si esta persona es apta para pagar el préstamo en un tiempo conveniente. Generalmente se piensa que, mientras más ingresos tenga la persona, es más probable que le acepten el crédito a financiar. Añadido a esto, también es frecuente escuchar que mientras la persona tenga un mayor grado académico, ésta tendrá un mejor salario o tendrá más formas de ingresar dinero a sus cuentas. Ante esta situación, nos surgió la idea de comprobar si estos estereotipos son ciertos ó no. Por lo que nuestra meta es lograr determinar si realmente el nivel de educación y el nivel de ingresos influye en la capacidad de pago de una persona en un préstamo, así como analizar si los factores cualitativos tienen cierta relevancia al momento de que haya un impago.

### 1.1.2 Conceptualización de la idea

“Verificar una relación entre la calificación de riesgo con las variables Nivel de educación, Ingresos, Monto del préstamo”. De la pregunta anterior, nos interesa conceptualizar dicha idea, por lo que, buscando la definición de las palabras que conforman la idea, en la RAE, encontramos lo siguiente:

- Relación: Conexión, correspondencia.
- Nivel: Medida de una cantidad con referencia a una escala determinada.
- Educación: Acción y efecto de educar, instrucción por medio de la acción docente.
- Ingresos: Caudal que entra en poder de alguien, y que le es de cargo en las cuentas.
- Calificación: Puntuación obtenida en un examen o en cualquier tipo de prueba.
- Riesgo: Contingencia o proximidad de un daño.

### 1.1.3 Identificación de tensiones

Como mencionamos en el apartado anterior, el trabajo se centra más en el estudio de las herramientas utilizadas, que en la información del trabajo per se. Sin embargo, vamos a desarrollar la teoría de manera satisfactoria con el hecho de crear un trabajo bien estructurado. Dicho lo anterior, es claro que la calificación de riesgo es un fenómeno que depende de muchas más variables, un ejemplo de ello puede ser la cultura de la sociedad en la cual se ve inmersa la persona que solicita el préstamo, como la entidad que lo desembolsa. Por ello, el origen de los datos de esta tabla de datos es de suma importancia, pues la cultura de las personas que arrojaron estos datos puede influir de manera sustancial en la salida de los datos.

Otro factor a tomar en cuenta es por supuesto, la tabla de datos, pues podría contener información errónea, o datos que no estén correctamente digitados, además de que la tabla de datos tiene que ser convertida a información más trabajable, es decir, muchas de las variables son categóricas, por lo que una conversión a datos de diferente especie, podría provocar que haya errores en la nueva tabla que se va a manipular.

### 1.1.4 Reformulación de la idea en modo preguntas

La idea principal de este trabajo es: “Verificar una relación entre la calificación de riesgo con las variables Nivel de educación, Ingresos, Monto del préstamo”, por lo que para esta parte del trabajo, vamos a formular la idea de diferentes formas, esto con el fin de formular una pregunta de investigación que sea clara.

*¿Cuáles son las variables que influyen más en la calificación de riesgo? ¿Existe una correlación positiva entre las variables de interés y la calificación de riesgo? ¿Es la calificación de riesgo un buen calificador de las variables que la determinan? ¿Cómo se mide la calificación de riesgo dadas las variables consideradas?*

### 1.1.5 Argumentación de las preguntas

Estas preguntas deben poseer una argumentación detrás de ellas, pues es necesario tener una idea de qué es lo que queremos realizar con estas preguntas o dicho de otra forma, cuál es el problema que potencialmente podemos resolver o la incógnita a contestar.<sup>8</sup>

#### 1.1.5.1 ¿Cuáles son las variables que influyen más en la calificación de riesgo?

##### Contraargumentos

Las variables que se toman son realmente las correctas, cuál es el método que se utiliza para determinar que las variables que se están tomando son mejores para un resultado que otras. Al ser una calificación de riesgo, que facilita el préstamo a personas o no, las personas podrían

tener intenciones perversas a siempre manipular su información para obtener buenas calificaciones. Por otro lado, son muchas las variables contempladas en la base de datos, hacer uso de un par de variables, podría influir de manera considerable en buscar una relación entre las variables.

### **Argumentos**

Al realizar un análisis previo de la tabla de datos, podríamos determinar cuáles son las variables con mayor impacto en esta calificación, y así simplificar el modelo y darle un mayor énfasis en estas variables que afectan de manera considerable en la calificación.

### **Concluya**

Al ser una calificación de riesgo, las empresas deben considerar varios factores, para ver si sus clientes son ideales para el préstamo o no. Por otro lado, concentrarnos en las variables que tienen mayor peso, podría ayudar a un mejor entendimiento de la calificación de riesgo.

#### **1.1.5.2 ¿Existe una correlación positiva entre las variables de interés y la calificación de riesgo?**

**Contraargumentos** Una relación positiva no implica que haya una mejor calificación, pues pueden existir variables con relación negativa, que tengan un efecto positivo sobre la calificación de riesgo, por pura intuición, podemos pensar en las personas que tienen un récord crediticio limpio, esto implicaría que esta variable debería tener un valor nulo para mejorar la calificación de riesgo. Otra variable que podría afectar es las veces que la persona ha caído en impago, pues entre más suba este valor, la calificación de riesgo, debería ser menor.

### **Argumentos**

Las relaciones que buscamos no tienen por qué ser positivas, desde un tipo de vista de coeficiente de correlación, sino desde un tipo de vista de mejoría, es decir, variables que en ausencia contribuyan a una mejor calificación, nos son de interés para el trabajo. Como hemos mencionado encontrar las relaciones más contribuyentes, son de utilidad, pues favorecen a simplificar el modelo y obtener a cambio una mejor interpretación del estudio.

### **Concluya**

Todas las relaciones son de importancia, hasta donde no existe relación, pues sirven para delimitar el modelo y determinar cuáles son las verdaderas variables que si influyen.

### **1.1.5.3 ¿Es la calificación de riesgo un buen calificador de las variables que la determinan?**

#### **Contraargumentos**

Esta pregunta, no es del todo objetiva, pues dependerá de lo que la empresa quiera detectar en estas evaluaciones, es decir, un factor que determina en gran medida a la calificación de riesgo, en otra empresa no tiene por qué ser así, pues dependerá del público objetivo de la empresa. Un buen calificador de riesgo depende tanto de las variables que se toman en cuenta, como del contexto de la empresa.

#### **Argumentos**

Desde el punto de vista de la empresa, la calificación de riesgo es una herramienta que sirve para determinar en gran medida si un préstamo se realiza o no, por ello, saber si su calificación de riesgo logra captar la información deseada de las variables utilizadas, entonces se podría considerar un buen factor, pues, aunque las empresas deban tener cuidado a quiénes otorgan los préstamos, también es un hecho, que si no lo hacen, se quedan sin negocio

#### **Concluya**

Si la calificación de riesgo logra simplificar y consolidar la información, entonces podríamos conjeturar que se comporta como una buena calificación.

### **1.1.5.4 ¿Cómo se mide la calificación de riesgo dadas las variables consideradas?**

#### **Contraargumentos**

La pregunta es más compleja que las hechas anteriormente, pues estamos entrando a un método de calificación, es decir, hacer un análisis de la medición, como mencionamos como un contraargumento en la pregunta anterior, esto no tiene por qué ser universal en las empresas, esto puede variar, por lo que dependerá del contexto en el cual se realice la pregunta.

#### **Argumentos**

En realidad que sea una pregunta que depende de sus variables, podría ser beneficioso, pues si existen varias metodologías, podríamos tener una mejor cartera de préstamos, es decir, dado cierto tipo de cliente, se podrían realizar cierto tipo de préstamos.

#### **Concluya**

La subjetividad de esta pregunta, no es ni buena ni mala, esta depende su contexto, si es aplicada de una buena forma, podría ser beneficioso tanto para la empresa, por ampliar su mercado, como para el cliente, al recibir el préstamo deseado.



### 1.1.6 Argumentación a través de datos

La base de datos utilizada en este trabajo se toma de la base de datos de Kaggle. Sin embargo, el autor de la base de datos no publica la información de cuándo es sacada la información. La información fue de fácil acceso pues está disponible en la de datos de Kaggle. La muestra observada son personas entre las edades de 18 a 69 años, de donde toman muchas variables, las cuales veremos más adelante. La unidad estadística estudiada para este trabajo son individuos que buscan obtener un préstamo.

Para la siguiente vamos a tomar el nombre las variables, las cuales dejaremos en el idioma original y vamos a dar la descripción de ellas, las cuales viene a su vez con la tabla de datos.

- Age: La edad del individuo, una variable continua que influye en la estabilidad financiera.
- Gender: Género del individuo, categorizado en Masculino, Femenino y No binario.
- Education Level: Nivel de educación alcanzado, que varía desde la Secundaria hasta el Doctorado.
- Marital Status: Estado civil actual, categorizado como Soltero, Casado, Divorciado o Viudo.
- Income: Ingreso anual en USD, que representa la capacidad de ganancia del individuo.
- Credit Score: Valor numérico que indica la solvencia crediticia, que varía de 600 a 800.
- Loan Amount: La cantidad de préstamo solicitada por el individuo, que representa las necesidades financieras.
- Loan Purpose: El propósito del préstamo, categorizado en Vivienda, Auto, Personal o Negocios.
- Employment Status: Situación laboral del individuo, incluyendo Empleado, Desempleado o Autónomo.
- Years at Current Job: Duración del empleo en el trabajo actual, que refleja la estabilidad laboral.
- Payment History: Desempeño histórico de pagos, categorizado como Excelente, Bueno, Regular o Malo.
- Debt-to-Income Ratio: Relación entre deuda e ingreso, que indica el apalancamiento financiero y el riesgo.
- Assets Value: Valor total de los activos poseídos por el individuo.
- Number of Dependents: Número de dependientes a cargo del individuo, que afecta las responsabilidades financieras.
- City: Ciudad en la que reside el individuo, proporcionando contexto geográfico.

- State: Estado en el que reside el individuo, proporcionando más detalles geográficos.
- Country: País de residencia, añadiendo una perspectiva global.
- Previous Defaults: Número de incumplimientos de préstamos anteriores, indicando el riesgo financiero histórico.
- Marital Status: Número de cambios en el estado civil, reflejando cambios en la vida personal.
- Risk Rating: Columna objetivo que categoriza el riesgo financiero en Bajo, Medio o Alto.

## **1.2 Revisión bibliográfica**

### **1.2.1 Búsqueda de bibliografía**

Entre las posibles combinaciones de palabras clave se se pueden encontrar:

- Ingreso + situación laboral + solvencia crediticia
- Prestamo + incumplimiento de préstamos + riesgo financiero
- Ciudad + propósito del prestamo + ingreso
- Situacion laboral + duración de empleo + historico de pagos
- Historico de pagos + incumplimiento de prestamos + categoria de riesgo
- Valor de activos + prestamo + relación entre deuda e ingreso

### **1.2.2 Fichas de literatura**

**Título: La valoración del riesgo financiero.**

- Autor: Dorina Chicu.
- Año: 2020.
- Nombre del tema: Métodos para medir el riesgo.
- Cronología: 2020.
- Metodología: Recolección de datos.
- Temática: Estudios económicos.
- Teórica: Valoración de riesgos.
- Resumen en una oración: Distintos riesgos existentes y algunas formas de medirlos.

- Argumento central: Analizar algunos de los distintos métodos existentes para medir los riesgos financieros.
- Problema con el argumento o el tema: Aunque el tema principal gire en torno a la valoración de riesgos financieros, el trabajo queda falente de varios detalles que, si pudieran ser notorios a la hora de hacer un análisis más exhaustivo, además de la falta de ejemplos u aplicaciones de estos, quedan solo como algo teórico.
- Resumen en un párrafo: El estudio busca centrarse en uno de los tres componentes de la inversión el cual es el riesgo financiero, de manera que se sabe que el objetivo principal de una empresa es maximizar sus beneficios, tal que llegue a asegurar la máxima rentabilidad posible. Por tanto, lo que se quiere brindar son los distintos métodos existentes que sirven para medir o valorar el riesgo, lo cual lleva a que sea posible generar una estrategia que permita mitigar los mismos.

**Título: La evaluación del riesgo de crédito en las instituciones de microfinanzas: estado del arte.**

- Autor: María Seijas, Milagro Vivel, Rubén Lado, Sara Fernández.
- Año: 2017.
- Nombre del tema: Riesgos en los microcréditos.
- Cronología: 2015 - 2017.
- Metodología: Recolección y comparación de datos.
- Temática: Estudios económicos.
- Teórica: Valoración de riesgos.
- Resumen en una oración: medición del riesgo de los microcréditos y análisis de los clientes.
- Argumento central: Explorar la diversa teoría existente al riesgo de crédito de las instituciones financieras.
- Problema con el argumento o el tema: Aunque el tema principal gire en torno a la valoración de riesgos financieros, el trabajo queda falente de varios detalles que, si pudieran ser notorios a la hora de hacer un análisis más exhaustivo, además de la falta de ejemplos u aplicaciones de estos, quedan solo como algo teórico.
- Resumen en un párrafo: Este trabajo busca exponer a través de las investigaciones que se han centrado en la evaluación del riesgo de crédito en las Instituciones de Microfinanzas, aquella teoría relacionada con el riesgo presente en los microcréditos, además busca analizar todos aquellos factores que llegan a ser determinantes en el riesgo de que haya algún tipo de impago., por lo que este estudio también ofrece ciertas técnicas que generan una mayor consistencia y transparencia en la evaluación y seguimiento de los clientes y su perfil.

**Título: Modelos para otorgamiento y seguimiento en la gestión del riesgo de crédito.**

- Autor: Millán Solarte, Julio César; Cerezo, Edinson Caicedo.
- Año: 2018.
- Nombre del tema: Riesgo de crédito.
- Cronología: 2018.
- Metodología: Análisis cuantitativa.
- Temática: Estudios económicos.
- Teórica: Valoración de riesgos.
- Resumen en una oración: Riesgos financieros y análisis en la gestión del riesgo de crédito.
- Argumento central: Explorar maneras de gestionar el riesgo financiero a través del análisis de las solicitudes.
- Problema con el argumento o el tema: Uno de los posibles problemas que se destacan en este estudio es que no se llega a profundizar en las posibles limitaciones que se pueden llegar a presentar a la hora de usar las técnicas que se narran y exploran, así como no se toma en cuenta que varias de las variables llegan a ser muy volátiles a lo largo del tiempo, además de los comportamientos de mercado que puedan incidir en el incumplimiento de pago.
- Resumen en un párrafo: Este trabajo habla acerca del riesgo financiero, más específicamente en el riesgo de crédito el cual se refiere a las perdidas derivadas del incumplimiento de obligaciones financieras, de manera que las instituciones financieras buscan gestionar este tipo de riesgos a través del análisis de las solicitudes por medio del sistema de scoring de crédito, dónde se evalúan variables como la situación financiera e historial de pagos del solicitante para diferenciar entre los buenos y malos clientes.

**Título: Variables determinantes de la probabilidad de incumplimiento de un microcrédito en una entidad microfinanciera del Perú, una aproximación bajo el modelo de regresión logística binaria.**

- Autor: María Calixto, Luis Casaverde
- Año: 2011.
- Nombre del tema: Incumplimiento de un microcrédito.
- Cronología: 2011.
- Metodología: Recolección y comparación de datos.
- Temática: Estudios económicos.

- Teórica: Valoración de riesgos.
- Resumen en una oración: Medición del riesgo de los microcréditos y el incumplimiento de pago.
- Argumento central: Explorar la probabilidad de incumplimiento de pago de los clientes a través del modelo de la regresión logística binaria.
- Problema con el argumento o el tema: La mayor problemática es que solo se enfoca en el modelo predictivo, de manera que no se llegan a abordar las posibles causas estructurales que al final terminan por inducir que hayan morosidades, de tal forma que los modelos como lo son la regresión logística binaria se va a ver muy limitada para lograr el propósito de reducir las morosidades.
- Resumen en un párrafo: Esta investigación gira entorno a los microcréditos y el incumplimiento de pago en estos, de manera que se busca identificar los factores que influyen en que exista mayor probabilidad de incumplimiento de los clientes, donde se utilizan modelos predictivos que permiten evaluar tanto variables cuantitativas como cualitativas relacionadas con las personas prestatarias, para ello se toman en cuenta factores socio-demográficos, económicos y financieros que lleguen a ser determinantes para el incumplimiento de pago.

**Título: Credit Scoring en Costa Rica y la probabilidad de clasificación de créditos personales basados en un modelo estadístico-matemático para aprobar o rechazar.**

- Autor: Patricia Hernández, Pablo Montoya, Allan Villareal
- Año: 2013.
- Nombre del tema: Incumplimiento de un microcrédito.
- Cronología: 2012-2013.
- Metodología: Recolección y comparación de datos.
- Temática: Estudios económicos.
- Teórica: Valoración de riesgos.
- Resumen en una oración: Analizar los riesgos en la designación de créditos a través de los modelos de credit scoring, tomando en cuenta variables cualitativas.
- Argumento central: Explorar la probabilidad de incumplimiento de pago de los clientes a través del modelo de la regresión logística binaria.

- Problema con el argumento o el tema: La mayor problemática es que solo se basa en los créditos personales para el consumo, y no toma en cuenta otros propósitos que se les puedan dar a los mismos, además de que en el estudio se menciona que no pudieron tener acceso a ciertas políticas que toman en cuenta los bancos para aceptar o rechazar un crédito.
- Resumen en un párrafo: Esta investigación se basa en el análisis de riesgos en la designación de créditos y el incumplimiento de pago a través de los modelos de credit scoring , de manera que analizar como a partir de estos se puede tratar de llegar a optimizar la asignación de los créditos y minimizar los riesgos, de manera que se termina analizando no solo el nivel de ingreso de una persona, sino también otros factores cualitativos como son la edad, el estado civil, el género, la escolaridad, entre otros.

## 1.3 Construcción de la UVE de Gowin

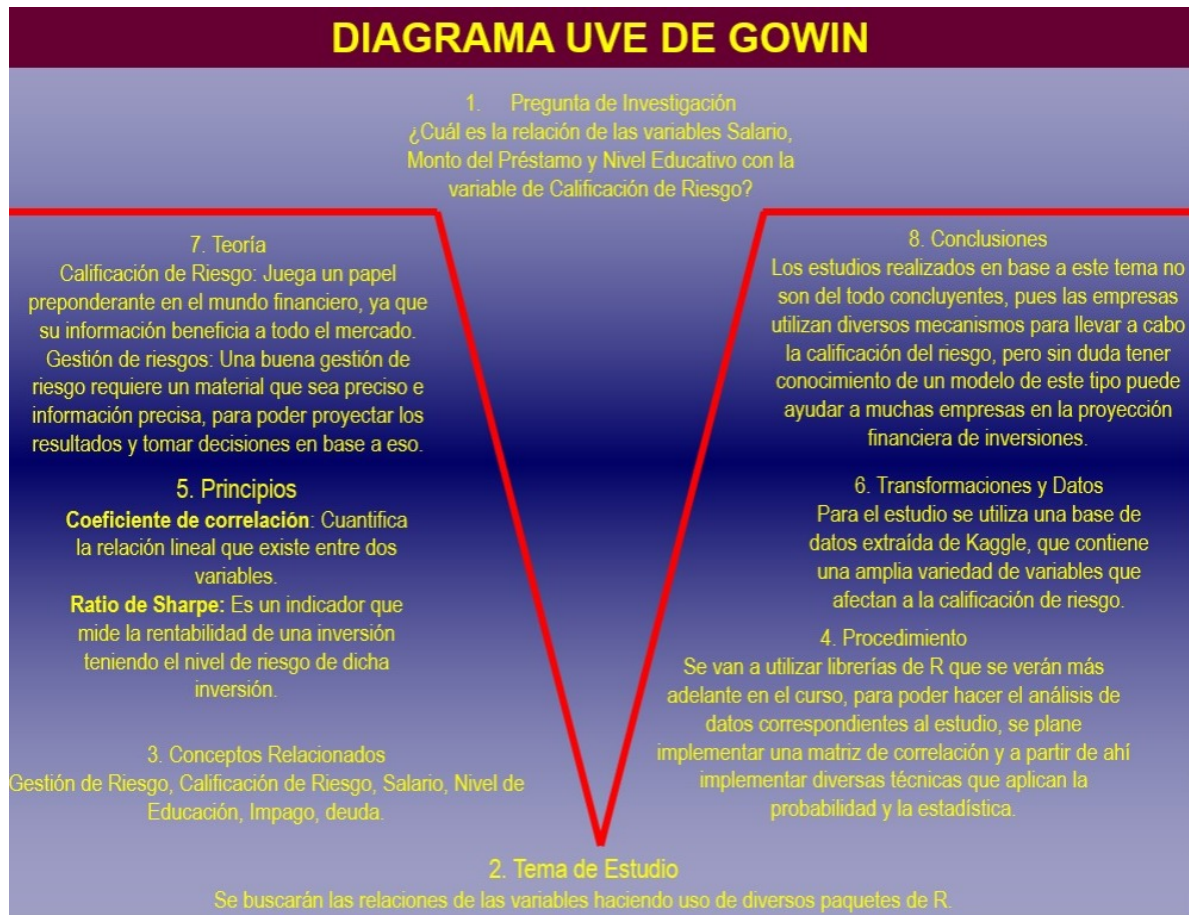


Figura 1.1: V de Gowin

### 1.3.1 Conceptos básicos

Como hemos estado hablando a lo largo del trabajo, la calificación de riesgo es de suma importancia, por lo que el estudio posee como objetivo determinar una relación fuerte de las variables, para aproximar un método empírico de una calificación de riesgo, pues ya vimos que las empresas por lo general no comparten esta información.

Por otro lado, la gestión de riesgo y la incertidumbre se vuelven esenciales a lo largo del trabajo, pues son conceptos que son base para el estudio posterior. El riesgo de impago también se vuelve un concepto a tener en cuenta, el cual vamos a entender como la incapacidad de cumplir una obligación financiera.

### 1.3.2 Principios y teorías

Para llevar a cabo la investigación en estas etapas preliminares hemos hecho estudio bibliográfico de lo que son los análisis de calificación financiera, que es justo el tema y variable objetivo de nuestro estudio. También se incorpora la teoría de la gestión de riesgos, que a su vez hace uso de técnicas de probabilidad pues dentro de su corpus la idea principal es maximizar una ganancia en eventos que presentan incertidumbre, de ahí la estrecha relación que tiene con Probabilidad.

Por otro lado, utilizar los conceptos de la valoración de riesgos financieros, sirve de base para entender cómo funciona el mercado financiero, y cuáles son los impactos directos de poseer una calificación de riesgos, y que el impacto es de hecho, para todos los participantes.

## 1.4 Parte de escritura

El problema que se va a tratar en el presente trabajo es de determinar la relación existente entre las variables de ingresos, nivel de educación y Monto del préstamo con la calificación de riesgo. Desde el punto de vista teórico, el autor (Palacios 2012), menciona que “La principal función que radica en las calificaciones crediticias es la evaluación de la mayor o menor probabilidad de pago de la deuda y los intereses, proporcionando indicadores que sirvan de referencia a los inversores con el fin de que puedan tener conocimiento del riesgo crediticio de una forma simple y accesible”. Desde este punto de vista, hay un apoyo en la investigación que tratamos de realizar, pues la calificación de riesgo es de suma importancia en el mundo financiero. Este mismo autor menciona que “Su importancia deriva de su implantación dentro de la regulación, lo que afecta a todo el entramado institucional, y sectores clave de la sociedad como son el bancario y las agencias de seguros y reaseguros”, podemos ver entonces que la teoría respalda la importancia que hemos estado conjeturando en esta presente bitácora (Entiéndase conjeturando, porque aún no hemos realizado el análisis de la tabla de datos).

El estudio del análisis financiero es de suma importancia en la actualidad, ya que las transacciones de los flujos de dinero cada vez son mayores, es decir, vender deuda para obtener financiamiento en el corto plazo es una de las estrategias más aplicadas, por ello tanto inversores como prestatarios, según (Palacios 2012), “Los inversores hacen uso de las calificaciones crediticias como un indicador de la probabilidad de recuperar su dinero. Adicionalmente, los prestatarios pueden beneficiarse de tener calificada su deuda, con el objetivo de colocarla con mayor facilidad y eliminar las dudas que haya relación a ellos.” Por ello, ambas parten obtienen beneficio de que exista este rating en el mundo de la información financiera. Y desde el punto de vista del inversor, como menciona la autora (Chicu 2020), “...a la hora de analizar una inversión, debemos valorar la rentabilidad esperada, así como la liquidez que perdemos y el riesgo que estamos dispuestos a sumir”. Por lo tanto, poseer la información de rating es de suma utilidad, pues ayuda a los inversores a realizar mejores proyecciones. En adición, haciendo referencia a esta misma autora “...la gestión de riesgos tiene un lugar cada vez que un



inversor analiza e intenta cuantificar las pérdidas potenciales en una inversión y luego toma las medidas apropiadas, considerando sus objetivos de inversión y su tolerancia al riesgo.” Esto último viene de la mano con lo que son las proyecciones, pues le permite al inversionista hacer un mejor análisis y una gestión de riesgos adecuada, que podemos definir según Chicu como “El proceso de identificación, análisis e incorporación de la incertidumbre en las decisiones de inversión” (Chicu 2020). Reforzando lo que menciona Chicu, la autora (Maria de los Ángeles Herrera 2024), menciona en adición a la gestión de riesgos “el contexto de incertidumbre genera inevitablemente un riesgo, y es ahí cuando la institución financiera debe preservar su valor económico y la integridad de los recursos confiados por los depositantes y socios.” Y añadiendo la definición de esta misma autora tenemos que la gestión de riesgos es “la denominación que se utiliza para el conjunto de técnicas y herramientas que pretenden maximizar el valor económico de la institución financiera, en un contexto de incertidumbre”. Concluyendo, la gestión de riesgos depende íntimamente de la calificación de riesgo, pues permite tener un parámetro ante la incertidumbre que representa invertir.

## 2 Bitácora 2

### 2.1 Parte de Planificación

#### 2.1.1 Ordenamiento de la Literatura

Tabla de Organización y Literatura

Tipo	Tema General	Tema Específico	Título	Año	Autor(es)
Metodológica	Correlación y Análisis de datos	Análisis del perfil de los prestatarios	La evaluación del riesgo de crédito en las instituciones de microfinanzas: estado del arte	2017	María Seijas, Milagro Vivel, Rubén Lado, Sara Fernández
Metodológica	Correlación y Análisis de datos	Relevancia del scoring de crédito	Modelos para el otorgamiento y seguimiento en la gestión del riesgo crediticio	2018	Millan Solarte, Julio Cesar, Edinson Caicedo
Metodológica	Correlación y Análisis de datos	Relación variables cuantitativas y cualitativas en el incumplimiento de pago	Variables determinantes de la probabilidad de incumplimiento de un microcrédito	2011	María Calixto, Luis Casaverde
Metodológica	Correlación y Análisis de datos	Clasificación de perfiles basándose en variables cualitativas	Credit Scoring en Costa Rica y la probabilidad de clasificación de créditos personales	2013	Patricia Hernández, José Montoya, Allan Villareal

```

if (!requireNamespace("kableExtra", quietly = TRUE)) {
  install.packages("kableExtra")
}
library(kableExtra)

data <- data.frame(
  Tipo = rep("Metodológica", 4),
  Tema_General = rep("Correlación y Análisis de datos", 4),
  Tema_Especifico = c("Análisis del perfil de los prestatarios",
    "Relevancia del scoring de crédito",
    "Relación variables cuantitativas y cualitativas en el incumplimiento",
    "Clasificación de perfiles basándose en variables cualitativas"),
  Título = c("La evaluación del riesgo de crédito en las instituciones de microfinanzas: estado del arte",
    "Modelos para el otorgamiento y seguimiento en la gestión del riesgo crediticio",
    "Variables determinantes de la probabilidad de incumplimiento de un microcrédito",
    "Credit Scoring en Costa Rica y la probabilidad de clasificación de créditos por riesgo"),
  Año = c(2017, 2018, 2011, 2013),
  Autor = c("María Seijas, Milagro Vivel, Rubén Lado, Sara Fernández",
    "Millan Solarte, Julio Cesar, Edinson Caicedo",
    "María Calixto, Luis Casaverde",
    "Patricia Hernández, José Montoya, Allan Villareal")
)

kable(data, format = "html", escape = FALSE,
  col.names = c("Tipo", "Tema General", "Tema Específico", "Título", "Año", "Autor(es)"),
  kable_styling(full_width = FALSE, position = "left")

```

## 2.2 Enlaces de la Literatura

**Título:** *La valoración del riesgo financiero.*

**Resumen:** El riesgo financiero es uno de los tipos de riesgo más importantes, por eso es de suma importancia entender su significado y qué herramientas se pueden utilizar para su debida gestión. En este caso se enfocan en tres métodos distintos: la desviación estándar, la beta del mercado y el valor en riesgo; así como comprender la rentabilidad ajustada con el ratio de Sharpe. Al utilizarlos en estrategias de cobertura de riesgo financiero se logra una reducción o mitigación del mismo en diferentes instrumentos financieros como las carteras de inversión.

**Contraste:** Este trabajo, a comparación del artículo de Solarte J. y Caicedo E. titulado Modelos para otorgamiento y seguimiento en la gestión del riesgo de crédito, logra explicar de

manera más amigable y general los conceptos del riesgo financiero, lo que hace que el artículo de Chicu sea más comprensible para todas las personas. Sin embargo, esta falta de rigurosidad puede generar fallas en los procesos más complejos del tema de investigación, ya que el mismo artículo de Solarte y Caicedo utiliza fórmulas y métodos más complejos que si bien no son tan fáciles de entender, logran una mejor recopilación de resultados.

**Comentario Propio:** Si bien nos parece importante que una explicación sea comprensible para la mayor cantidad de personas, no tener la suficiente precisión en un tema tan complicado como lo es el riesgo financiero puede causar ignorancia en los conceptos avanzados, como por ejemplo los métodos paramétricos y no paramétricos y su metodología. Sin embargo, sentimos que esta investigación nos ayudó con la afinidad a la hora de seleccionar las palabras adecuadas en el desarrollo de nuestra investigación.

**Título:** *La evaluación del riesgo de crédito en las instituciones de microfinanzas: estado del arte.*

**Resumen:** El análisis de los modelos de credit scoring muestra que estos pueden ser utilizados en diversas dimensiones, aunque hay una preferencia por predecir el riesgo de retrasos costosos en los microcréditos los cuales están bajo control de las instituciones financieras para mitigarlos. La literatura teórica destaca la importancia de la información estadística cualitativa en estos modelos, además de incluir variables relacionadas con el prestatario, su negocio y el préstamo. Existe una preferencia por las técnicas paramétricas como por ejemplo la regresión logística. Aún así, en los últimos años se registra un aumento en el uso de técnicas no paramétricas que logran un mayor poder predictivo en la detección de incumplimientos de microcréditos. Con esto, las instituciones están incorporando estas herramientas para convertirlas en una práctica estándar, como en las instituciones bancarias; las cuales mejoran la consistencia, transparencia, control de calidad y optimización de los procesos.

**Contraste:** Éste trabajo, a diferencia del artículo de Chicu D. titulado La valoración del riesgo financiero, logra realizar no sólo un análisis del credit scoring por medio de teoría estadística, sino que también logra plantear un análisis macroeconómico y toma en cuenta las repercusiones que las fallas en los procesos de gestión de riesgos podría generar en el sistema financiero. Sin embargo, no es tan específico en la parte matemática a la hora de estimar probabilidades de que algo pase.

**Comentario Propio:** Para nosotros es importante entender también la teoría desde la perspectiva economista ya que nos ayuda a entender a qué nos enfrentamos y cómo podemos aplicar diferentes metodologías de investigación.

**Título:** *Modelos para otorgamiento y seguimiento en la gestión del riesgo de crédito.*

**Resumen:** El credit scoring es un método estadístico para estimar la probabilidad de incumplimiento de un prestatario, usando su información histórica y estadística para obtener un indicador que permita distinguir la calidad de un deudor. Los modelos de scoring son muy importantes para los procesos de gestión del crédito, los cuales buscan explicar la composición y operatividad de estos modelos utilizando grandes bases de datos. La información que resulta

de estos modelos permite el análisis de la toma de decisión de si se otorga ó no un crédito a una persona. Por medio de cuatro modelos distintos se logra un procedimiento multicapa para obtener información precisa para calificar a un cliente como bueno o malo para la empresa financiera.

**Contraste:** Este trabajo logra ser más específico que los demás en cuanto a los métodos de gestión de riesgo se refiere. Además coincide con varios de los artículos recopilados respecto a la utilización del método de regresión logística como herramienta de calificación de crédito.

**Comentario Propio:** Consideramos que este artículo es el más completo en cuanto a materia matemática se refiere, ya que logra explicar los métodos de manera clara y además los muestra con las gráficas y resultados obtenidos. De hecho, es de gran ayuda que las variables a trabajar sean similares a las que la investigación utiliza.

**Título:** *Variables determinantes de la probabilidad de incumplimiento de un microcrédito en una entidad microfinanciera del Perú, una aproximación bajo el modelo de regresión logística binaria.*

**Resumen:** Después de realizar una estimación probabilística del incumplimiento de un microcrédito por medio de la aproximación de una función logística binaria, se determina que son las variables cualitativas como el estado civil, edad y tipo de vivienda junto con las variables de plazo, número de créditos con la entidad y el saldo deudor; las que generan un modelo correctamente ajustado bajo el modelo de regresión logística. La cual logra una capacidad predictiva aceptable medida por la curva ROC.

**Contraste:** Este trabajo, a diferencia de “La evaluación del riesgo de crédito en las instituciones de microfinanzas: estado del arte”, determina las variables necesarias para un estudio de microfinanzas. Sin embargo, coinciden en todo lo referente a las ventajas y desventajas del uso de la calificación crediticia, aunque si cabe mencionar que esta investigación también logra hacer conclusiones de los inversores y reguladores como parte de los responsables de la ineficiencia que ha dado el uso excesivo de ésta calificación.

**Comentario Propio:** Fue de mucha ayuda entender la noción de qué variables utilizar y porqué son importantes para un estudio de calificación de crédito.

**Título:** *Calificación de riesgo: definición e influencia en la última década.*

**Resumen:** Las agencias de calificación crediticia han crecido en las últimas décadas por su capacidad para reducir las asimetrías de información en los mercados, facilitando la liquidez y aumentando los participantes. Sin embargo, no han cumplido con los efectos positivos esperados en la última década, se han expuesto fallas en su funcionamiento, como su papel en la burbuja de deuda y la crisis económica, lo que ha generado inestabilidad y ralentizado la recuperación. La importancia que los inversores le dieron a las calificaciones crediticias para la toma de decisiones fue desmedida al no tomar en cuenta el nivel de riesgo, lo cual se relaciona a las causas de las últimas crisis financieras e hipotecarias.

**Contraste:** En comparación de la investigación titulada “La evaluación del riesgo de crédito en las instituciones de microfinanzas: estado del arte”, ambos logran concluir cosas muy similares en cuanto a la importancia del credit score, sus ventajas y desventajas. Sin embargo, logra aportar más en cuanto a las ineficiencias que ha tenido sobrevalorarlo y los momentos en los que ha generado una crisis.

**Comentario Propio:** Este trabajo, de igual forma, es de suma importancia para entender la manera en que se realizan calificaciones crediticias y de qué formas analizarlas. Para así utilizarlas en nuestros estudios estadísticos y entender qué otras variables ó conceptos tomar en cuenta para los futuros resultados a obtener.

**Título:** Conceptualización del riesgo de los mercados financieros.

**Resumen:** El riesgo está inmerso en todas las actividades humanas y es entendido como la probabilidad de ocurrencia de un evento que podría inducir un perjuicio. Cuando se habla de riesgo financiero, se habla de una eventual pérdida de dinero que signifique una afectación al sistema financiero ó a alguna institución que sea parte del mismo. Este trabajo es una recopilación algunos de los riesgos en los mercados financieros y presenta algunos métodos válidos para su valoración.

**Contraste:** Esta investigación, a diferencia de las demás investigaciones, logra tener un glosario completo de definiciones relacionadas al riesgo financiero, sin embargo no va más allá de ser sólomente un artículo de definiciones. Por lo que en materia teórica no tiene nada que aportar.

**Comentario Propio:** Gracias a estas definiciones, hemos logrado definir mejor qué queremos estudiar de manera más específica y cómo utilizar ciertos conceptos de mejor manera.

## 2.3 Análisis Estadístico

Como base para realizar este análisis estadístico, nos estamos guiando con la guía del curso de Herramientas de Ciencias de Datos, el cual adjuntamos el link a dicha guía (Solis 2024) y también estamos utilizando el libro escrito por Wickham, el cual también adjuntamos el link (Hadley Wickham 2019).

A modo de introducción, el análisis estadístico consiste en un conjunto de herramientas o técnicas que se utilizan para la recolección, el análisis e interpretación de datos. Para este trabajo es imprescindible contar con este set de herramientas.

### 2.3.1 Análisis Descriptivo

La base de datos ya se encuentra en formato en tidy, recordemos que el formato tidy fue popularizado por el autor Hadley Wickham, donde indican que cada variable debe tener su

propia columna y cada observación su propia fila. Nuestra base de datos cumple con estar en formato tidy.

Vamos a llamar a nuestra base de datos, la cual vamos a utilizar durante el trabajo.

```
Rows: 15000 Columns: 20
```

```
-- Column specification -----
```

```
Delimiter: ";"
```

```
chr (10): Gender, Education Level, Marital Status, Loan Purpose, Employment ...
```

```
dbl (10): Age, Income, Credit Score, Loan Amount, Years at Current Job, Debt...
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(base_financial_risk_assessment)
```

```
# A tibble: 6 x 20
```

	Age	Gender	Education Level	Marital Status	Income	Credit Score
	<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>
1	49	Male	PhD	Divorced	72799	688
2	57	Female	Bachelor's	Widowed	NA	690
3	21	Non-binary	Master's	Single	55687	600
4	59	Male	Bachelor's	Single	26508	622
5	25	Non-binary	Bachelor's	Widowed	49427	766
6	30	Non-binary	PhD	Divorced	NA	717

```
# i 14 more variables: `Loan Amount` <dbl>, `Loan Purpose` <chr>,  
# `Employment Status` <chr>, `Years at Current Job` <dbl>,  
# `Payment History` <chr>, `Debt-to-Income Ratio` <dbl>,  
# `Assets Value` <dbl>, `Number of Dependents` <dbl>, City <chr>,  
# State <chr>, Country <chr>, `Previous Defaults` <dbl>,  
# `Marital Status Change` <dbl>, `Risk Rating` <chr>
```

Antes de aplicar cualquier gráfico o análisis de datos a nuestra base de datos, es importante eliminar las variables que no aportan al estudio, por ello, vamos a eliminar los valores NA que vengan en nuestra base de datos.

```
Base_limpia <- na.omit(base_financial_risk_assessment)  
head(Base_limpia)
```

```
# A tibble: 6 x 20
```

	Age	Gender	Education Level	Marital Status	Income	Credit Score
--	-----	--------	-----------------	----------------	--------	--------------

	<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>
1	49	Male	PhD	Divorced	72799	688
2	21	Non-binary	Master's	Single	55687	600
3	59	Male	Bachelor's	Single	26508	622
4	42	Non-binary	Master's	Single	116212	707
5	55	Male	High School	Married	70978	706
6	56	Non-binary	PhD	Married	21084	702

```
# i 14 more variables: `Loan Amount` <dbl>, `Loan Purpose` <chr>,
# `Employment Status` <chr>, `Years at Current Job` <dbl>,
# `Payment History` <chr>, `Debt-to-Income Ratio` <dbl>,
# `Assets Value` <dbl>, `Number of Dependents` <dbl>, City <chr>,
# State <chr>, Country <chr>, `Previous Defaults` <dbl>,
# `Marital Status Change` <dbl>, `Risk Rating` <chr>
```

Ahora hacemos un análisis estadístico de nuestra base de datos, de todas las variables.

```
# Instalar kableExtra si no lo tienes
# install.packages("kableExtra")

library(dplyr)
library(tidyr)
library(knitr)
library(kableExtra)

# Como nuestras variables no son del todo numéricas, hay que hacerlo para las variables que s

# Primero hacemos las variables numéricas.
resumen_numericas <- Base_limpia %>%
  summarise(
    Edad_Media = mean(Age, na.rm = TRUE),
    Edad_Minima = min(Age, na.rm = TRUE),
    Edad_Maxima = max(Age, na.rm = TRUE),
    Ingreso_Medio = mean(Income, na.rm = TRUE),
    Ingreso_Varianza = var(Income, na.rm = TRUE),
    Ingreso_Minimo = min(Income, na.rm = TRUE),
    Ingreso_Maximo = max(Income, na.rm = TRUE),
    Prestamo_Medio = mean(`Loan Amount`, na.rm = TRUE),
    Prestamo_Varianza = var(`Loan Amount`, na.rm = TRUE),
    Prestamo_Minimo = min(`Loan Amount`, na.rm = TRUE),
    Prestamo_Maximo = max(`Loan Amount`, na.rm = TRUE)
  )
```



```

# Variables Categóricas
resumen_categoricas <- tibble(
  Variable = c("Nivel de Educación", "Género", "Estado Civil"),
  Frecuencia = c(
    paste(names(table(Base_limpia$`Education Level`)), collapse = ", "),
    paste(names(table(Base_limpia$Gender)), collapse = ", "),
    paste(names(table(Base_limpia$`Marital Status`)), collapse = ", ")
  )
)

# Luego resumimos la información en un cuadro, para presentarlo mejor.
tabla_resumen <- data.frame(
  Variable = c("Edad", "Ingresos", "Monto del Préstamo", resumen_categoricas$Variable),
  Media = c(resumen_numericas$Edad_Media, resumen_numericas$Ingreso_Medio, resumen_numericas$Prestamo_Medio, resumen_categoricas$Frecuencia),
  Varianza = c(NA, resumen_numericas$Ingreso_Varianza, resumen_numericas$Prestamo_Varianza, NA),
  Mínimo = c(resumen_numericas$Edad_Minima, resumen_numericas$Ingreso_Minimo, resumen_numericas$Prestamo_Minimo, NA),
  Máximo = c(resumen_numericas$Edad_Maxima, resumen_numericas$Ingreso_Maximo, resumen_numericas$Prestamo_Maximo, NA),
  Frecuencia = c(rep(NA, 3), resumen_categoricas$Frecuencia)
)

# Mostrar la tabla con la descripción
tabla_resumen_kable <- kable(tabla_resumen, caption = "Resumen de Variables Numéricas y Categóricas",
  kable_styling() %>%
  add_header_above(c(" " = 1, "Resumen" = 5))
tabla_resumen_kable <- tabla_resumen_kable %>%
  kableExtra::footnote(general = "Fuente: Elaboración propia utilizando la base de datos de Kaggle")
tabla_resumen_kable

```

Tabla 2.2: Resumen de Variables Numéricas y Categóricas

Variable	Resumen				
	Media	Varianza	Mínimo	Máximo	Frecuencia
Edad	43.5817	NA	18	69	NA
Ingresos	70190.3585	849685113	20014	119978	NA
Monto del Préstamo	27577.0679	168240584	5001	49978	NA
Nivel de Educación	NA	NA	NA	NA	Bachelor's, High School, Master's, PhD
Género	NA	NA	NA	NA	Female, Male, Non-binary
Estado Civil	NA	NA	NA	NA	Divorced, Married, Single, Widowed

*Note:*

Fuente: Elaboración propia utilizando la base de datos de Kaggle

Presentamos las variables de más importancia en nuestro estudio

Tabla 2.3: Resumen de Variables Numéricas y Categóricas

Variable	Resumen				
	Media	Varianza	Mínimo	Máximo	Frecuencia
Edad	43.5817	NA	18	69	NA
Ingresos	70190.3585	849685113	20014	119978	NA
Monto del Préstamo	27577.0679	168240584	5001	49978	NA
Nivel de Educación	NA	NA	NA	NA	Bachelor's, High School, Master's, PhD
Género	NA	NA	NA	NA	Female, Male, Non-binary
Estado Civil	NA	NA	NA	NA	Divorced, Married, Single, Widowed

*Note:*

Fuente: Elaboración propia utilizando la base de datos de Kaggle

Con nuestra base de datos limpia, vamos a proceder a calcular algunos estadísticos importantes, por separado. Estos son los datos resumidos de la variable edad.

```
library(dplyr)

# Calculamos las estadísticas de interés para las variables y las almacenamos en un solo dato
resultado <- Base_limpia %>%
  summarise(
    # Estadísticos para la variable edad/Age
    media_edad = mean(Age, na.rm = TRUE),
    varianza_edad = var(Age, na.rm = TRUE),
    min_edad = min(Age, na.rm = TRUE),
    max_edad = max(Age, na.rm = TRUE),

    # Estadísticos para la variable ingreso/Income
    media_ingresos = mean(Income, na.rm = TRUE),
    varianza_ingresos = var(Income, na.rm = TRUE),
    min_ingresos = min(Income, na.rm = TRUE),
    max_ingresos = max(Income, na.rm = TRUE),

    # Estadísticos para la variable record_crediticio/Credit Score
    media_record_crediticio = mean(`Credit Score`, na.rm = TRUE),
    varianza_record_crediticio = var(`Credit Score`, na.rm = TRUE),
    min_record_crediticio = min(`Credit Score`, na.rm = TRUE),
    max_record_crediticio = max(`Credit Score`, na.rm = TRUE),
```

```

# Estadísticos para la variable monto del préstamo/Loan Amount
media_monto_prestamo = mean(`Loan Amount`, na.rm = TRUE),
varianza_monto_prestamo = var(`Loan Amount`, na.rm = TRUE),
min_monto_prestamo = min(`Loan Amount`, na.rm = TRUE),
max_monto_prestamo = max(`Loan Amount`, na.rm = TRUE),

# Estadísticos para la variable años de trabajo/Years at Current Job
media_anyos_trabajo = mean(`Years at Current Job`, na.rm = TRUE),
varianza_anyos_trabajo = var(`Years at Current Job`, na.rm = TRUE),
min_anyos_trabajo = min(`Years at Current Job`, na.rm = TRUE),
max_anyos_trabajo = max(`Years at Current Job`, na.rm = TRUE)
)

# Imprimimos el resultado
print(resultado)

# A tibble: 1 x 20
  media_edad varianza_edad min_edad max_edad media_ingresos varianza_ingresos
    <dbl>         <dbl>    <dbl>    <dbl>         <dbl>         <dbl>
1    43.6         218.      18      69      70190.      849685113.
# i 14 more variables: min_ingresos <dbl>, max_ingresos <dbl>,
#   media_record_crediticio <dbl>, varianza_record_crediticio <dbl>,
#   min_record_crediticio <dbl>, max_record_crediticio <dbl>,
#   media_monto_prestamo <dbl>, varianza_monto_prestamo <dbl>,
#   min_monto_prestamo <dbl>, max_monto_prestamo <dbl>,
#   media_anyos_trabajo <dbl>, varianza_anyos_trabajo <dbl>,
#   min_anyos_trabajo <dbl>, max_anyos_trabajo <dbl>

resultado$media_edad

[1] 43.5817

resultado$varianza_edad

[1] 217.7303

resultado$min_edad

[1] 18

```

```
resultado$max_edad
```

```
[1] 69
```

Estos son los datos resumidos de la variable Ingresos

```
resultado$media_ingresos
```

```
[1] 70190.36
```

```
resultado$varianza_ingresos
```

```
[1] 849685113
```

```
resultado$min_ingresos
```

```
[1] 20014
```

```
resultado$max_ingresos
```

```
[1] 119978
```

Estos son los datos resumidos de la variable Monto del Préstamo

```
resultado$media_monto_prestamo
```

```
[1] 27577.07
```

```
resultado$varianza_monto_prestamo
```

```
[1] 168240584
```

```
resultado$min_monto_prestamo
```

```
[1] 5001
```

```
resultado$max_monto_prestamo
```

```
[1] 49978
```

Resumimos la información obtenida en un cuadro, para una mejor visualización de ellos.

```
library(knitr)
library(kableExtra)

# Creamos una tabla para resumir la información obtenida
tabla_resumen <- data.frame(
  Variable = c("Edad", "Ingresos", "Monto del Préstamo"),
  Media = c(resultado$media_edad, resultado$media_ingresos, resultado$media_monto_prestamo),
  Varianza = c(NA, resultado$varianza_ingresos, resultado$varianza_monto_prestamo),
  Mínimo = c(resultado$min_edad, resultado$min_ingresos, resultado$min_monto_prestamo),
  Máximo = c(resultado$max_edad, resultado$max_ingresos, resultado$max_monto_prestamo)
)

# Mostramos la tabla con la descripción
tabla_resumen_kable <- kable(tabla_resumen, caption = "Resumen de Variables: Edad, Ingresos y Monto del Préstamo",
  kable_styling()

tabla_resumen_kable <- tabla_resumen_kable %>%
  kableExtra::footnote(general = "Fuente: Elaboración propia utilizando la base de datos de Kaggle")

tabla_resumen_kable
```

Tabla 2.4: Resumen de Variables: Edad, Ingresos y Monto del Préstamo

Variable	Media	Varianza	Mínimo	Máximo
Edad	43.5817	NA	18	69
Ingresos	70190.3585	849685113	20014	119978
Monto del Préstamo	27577.0679	168240584	5001	49978

*Note:*

Fuente: Elaboración propia utilizando la base de datos de Kaggle

Por último, vamos a realizar una matriz de correlación de los datos, esto porque queremos observar la relación que tienen las variables, solo tomaremos las variables de interés, la justificación de dicha escogencia viene del lado teórico, pues son las variables que históricamente más se toman en los estudios de calificación de riesgo.

```

library(dplyr)
library(corrplot)

# Escogemos las variables que nos interesan para la matriz de correlación.
Base_correlacion <- Base_limpia %>%
  select(`Risk Rating`, Income, `Loan Amount`, Age, `Loan Purpose`, `Education Level`)

# Como hay variables categóricas, entonces vamos a convertir las variables a numérico, para p
Base_correlacion$`Risk Rating` <- as.numeric(as.factor(Base_correlacion$`Risk Rating`))
Base_correlacion$`Loan Purpose` <- as.numeric(as.factor(Base_correlacion$`Loan Purpose`))
Base_correlacion$`Education Level` <- as.numeric(as.factor(Base_correlacion$`Education Level`))

# Calculamos la matriz de correlación
matriz_correlacion <- cor(Base_correlacion, use = "complete.obs", method = "pearson")
print(matriz_correlacion)

```

	Risk Rating	Income	Loan Amount	Age
Risk Rating	1.000000000	0.013528536	-0.015100412	0.003258428
Income	0.013528536	1.000000000	-0.008137282	0.005019572
Loan Amount	-0.015100412	-0.008137282	1.000000000	-0.011121494
Age	0.003258428	0.005019572	-0.011121494	1.000000000
Loan Purpose	-0.015622201	0.014753633	0.006311870	-0.013760821
Education Level	-0.013449909	0.019406630	0.010511349	0.011114696

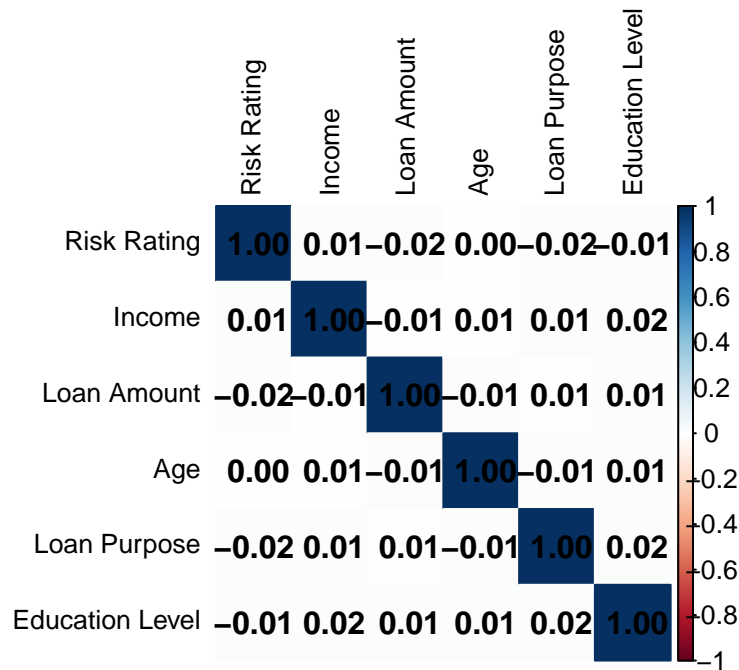
  

	Loan Purpose	Education Level
Risk Rating	-0.01562220	-0.01344991
Income	0.01475363	0.01940663
Loan Amount	0.00631187	0.01051135
Age	-0.01376082	0.01111470
Loan Purpose	1.00000000	0.01934448
Education Level	0.01934448	1.00000000

```

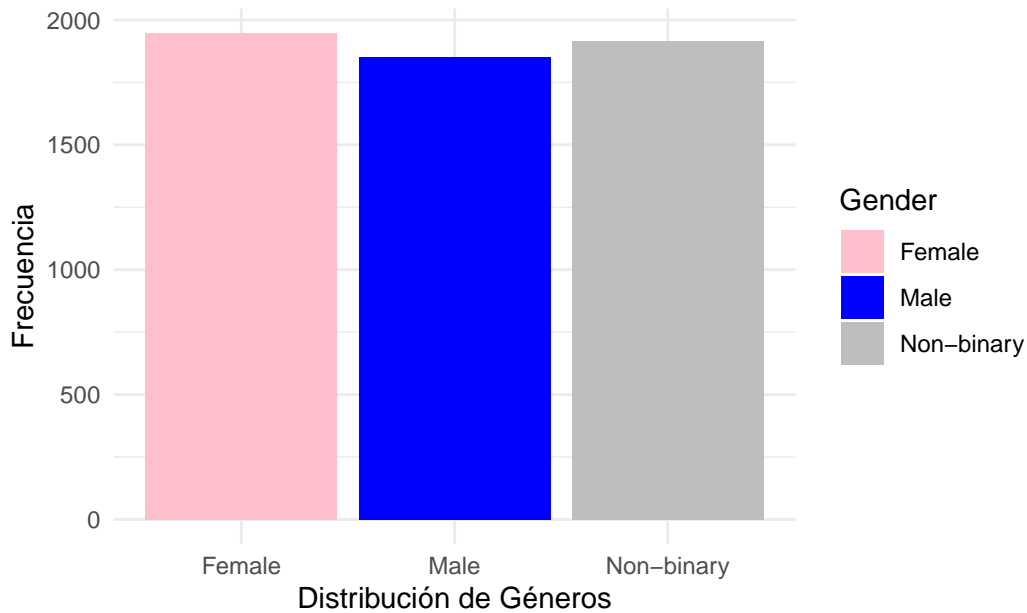
# Creamos el gráfico de la matriz de correlación
corrplot(matriz_correlacion, method = "color", addCoef.col = "black", tl.col = "black", tl.c
mtext("Fuente: Elaboración propia utilizando la base de datos de Kaggle", side = 1, line = 4

```



Ahora vamos a comenzar a ver los gráficos que se forman de nuestra base de datos. Para ello vamos a analizar cómo es la distribución que siguen los géneros de nuestra base de datos, esto es solamente por sondear cómo es nuestra población. Al ser una variable categórica, lo recomendado es realizar un gráfico de barras.

```
library(ggplot2)
# Gráfico con las distribuciones del género
ggplot(Base_limpia, aes(x = Gender, fill = Gender)) +
  geom_bar() +
  scale_fill_manual(values = c("Male" = "blue", "Female" = "pink", "Non-binary" = "gray")) +
  labs(x = "Distribución de Géneros", y = "Frecuencia") + theme_minimal() +
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

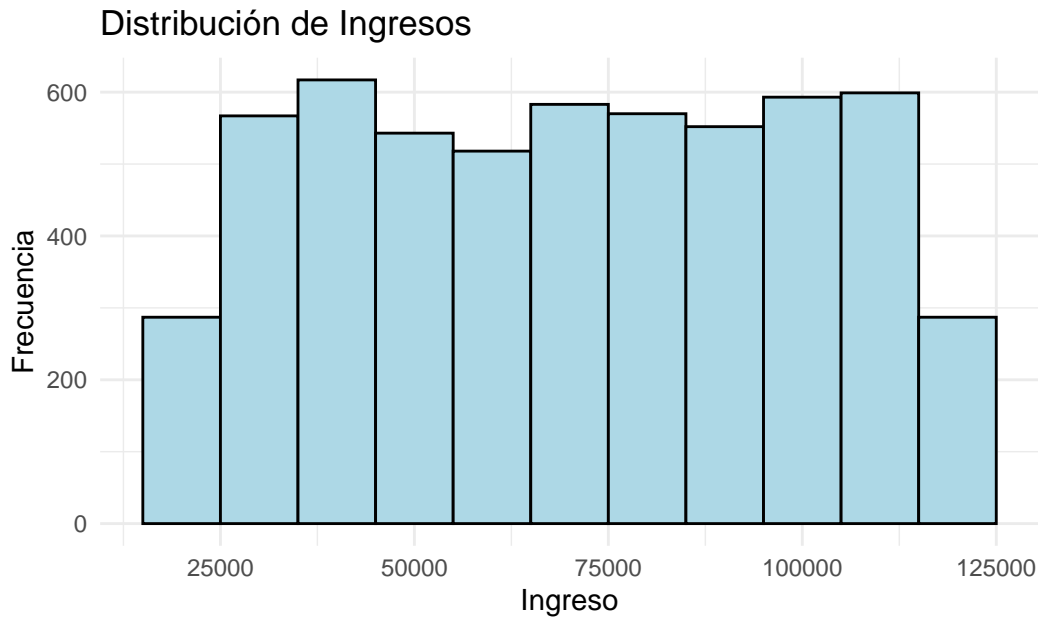


Fuente: Elaboración propia utilizando la base de datos de Kaggle

Por otro lado, la variable “Income”, es una variable continua, por lo que lo recomendado para visualizar la distribución de estos datos de una manera rápida es a través de los histogramas, por lo que adjuntamos el gráfico correspondiente:

```
library(ggplot2)
ggplot(Base_limpia, aes(x = Income)) +
  geom_histogram(binwidth = 10000, fill = "lightblue", color = "black") +
  labs(x = "Ingreso", y = "Frecuencia", title = "Distribución de Ingresos") +
  theme_minimal() +
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```



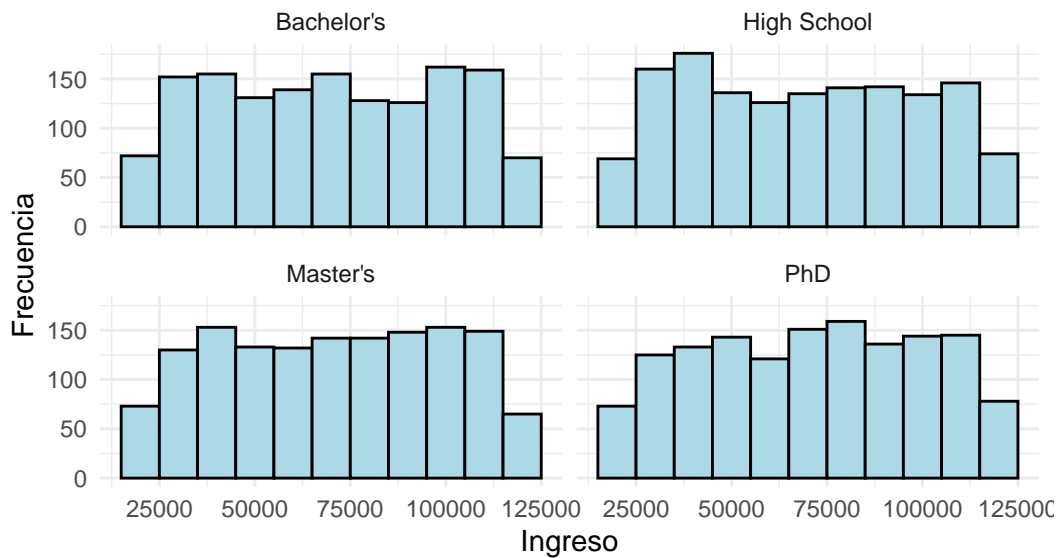


Fuente: Elaboración propia utilizando la base de datos de Kaggle

Además decidimos realizar un facet en esta misma variable con respecto a la variable “Education Level”, con el fin de visualizar la distribución en cada categoría, esto porque queremos descartar o validar que de alguna forma el nivel educativo tiene relación con nuestra variable objetivo, la cual es la calificación de riesgo.

```
library(ggplot2)
ggplot(Base_limpia, aes(x = Income)) +
  geom_histogram(binwidth = 10000, fill = "lightblue", color = "black") +
  labs(x = "Ingreso", y = "Frecuencia", title = "Distribución de Ingresos") + theme_minimal()
  facet_wrap(~ `Education Level`) + # Facet por nivel educativo
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

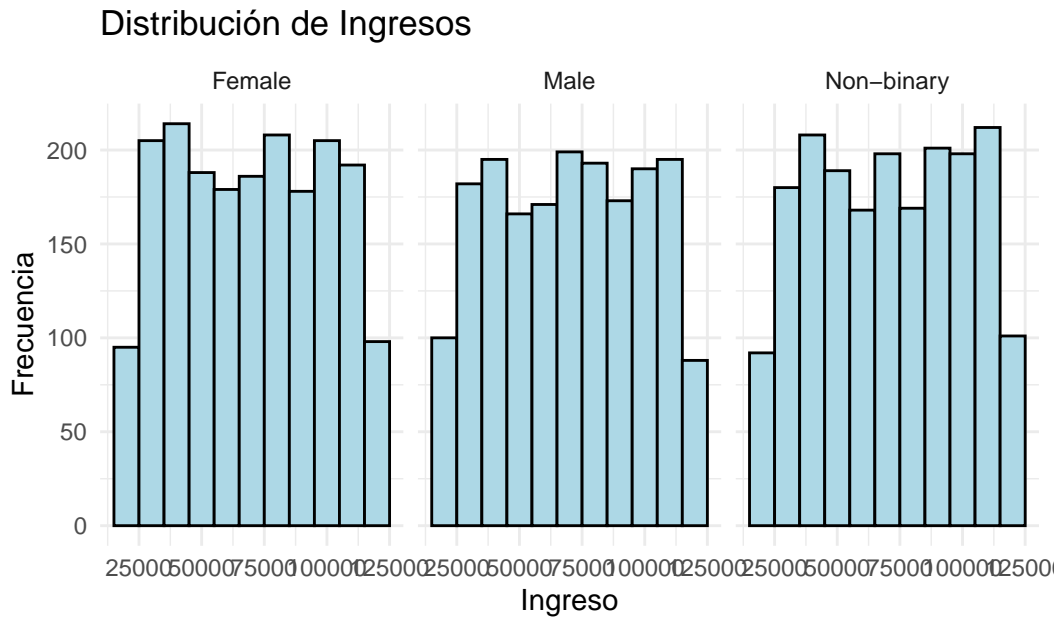
## Distribución de Ingresos



Fuente: Elaboración propia utilizando la base de datos de Kaggle

Adjuntamos un gráfico más de los ingresos, pero esta vez por género, con la misma curiosidad de observar cómo es la distribución de estos dada la mencionada variable.

```
library(ggplot2)
ggplot(Base_limpia, aes(x = Income)) +
  geom_histogram(binwidth = 10000, fill = "lightblue", color = "black") +
  labs(x = "Ingreso", y = "Frecuencia", title = "Distribución de Ingresos") + theme_minimal()
  facet_wrap(~ Gender) + # Facet Género
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

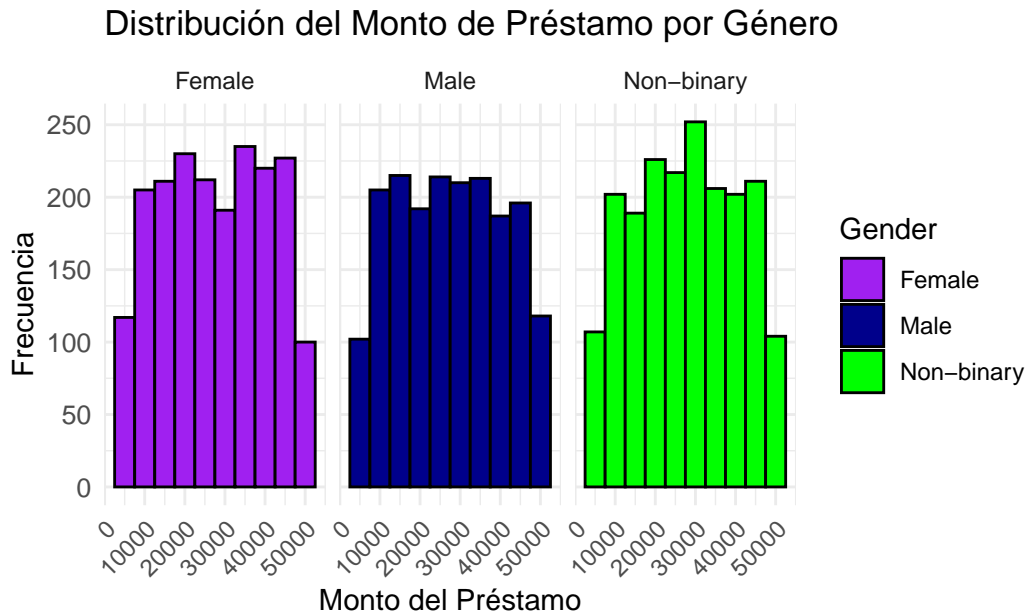


Fuente: Elaboración propia utilizando la base de datos de Kaggle

Otras variables a tener en cuenta aparte de las anteriores mencionadas con respecto al riesgo de crédito, son el monto del préstamos y el propósito del préstamo. Vamos analizar primero el Monto del préstamo, nos interesa ver qué distribución en general tiene.

```
library(ggplot2)

ggplot(Base_limpia, aes(x = `Loan Amount`, fill = Gender)) +
  geom_histogram(binwidth = 5000, color = "black", position = "identity") +
  scale_fill_manual(values = c("Male" = "darkblue", "Female" = "purple", "Non-binary" = "green")) +
  facet_wrap(~ Gender) +
  labs(x = "Monto del Préstamo", y = "Frecuencia", title = "Distribución del Monto de Préstamo") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10), # Ajustamos el tamaño del texto
        axis.text.x = element_text(angle = 45, hjust = 1)) + #Rotamos el texto
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```



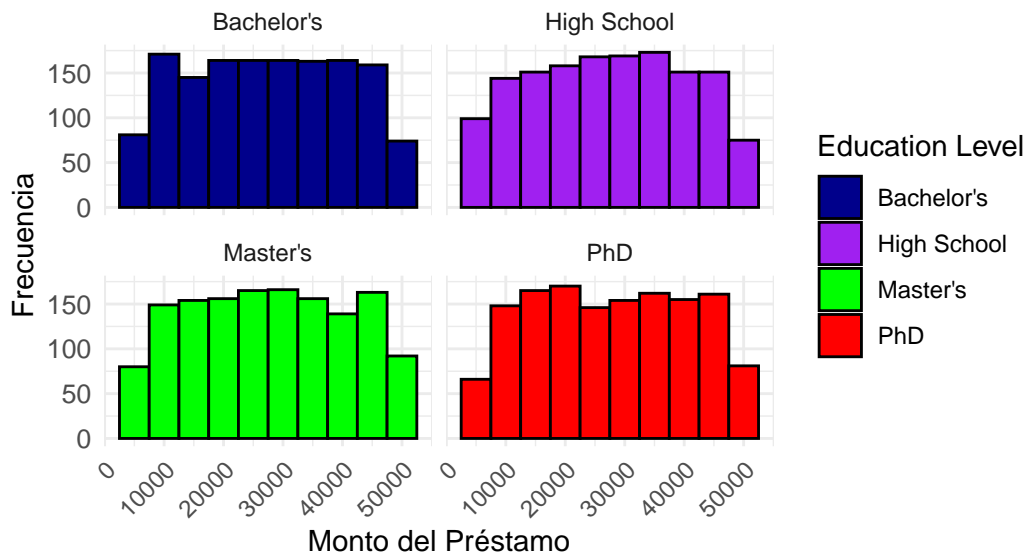
Fuente: Elaboración propia utilizando la base de datos de Kaggle

Haremos lo mismo, pero esta vez vamos a ver cómo se comportan los salarios cuando cambiamos la variable y utilizamos por ejemplo el grado académico.

```
library(ggplot2)

ggplot(Base_limpia, aes(x = `Loan Amount`, fill = `Education Level`)) +
  geom_histogram(binwidth = 5000, color = "black", position = "identity") +
  scale_fill_manual(values = c("Bachelor's" = "darkblue", "High School" = "purple", "Master's" = "green")) +
  facet_wrap(~ `Education Level`) +
  labs(x = "Monto del Préstamo", y = "Frecuencia", title = "Distribución del Monto de Préstamo por Nivel Educativo") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10), # Ajustamos el tamaño del texto
        axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotamos el texto
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

## Distribución del Monto de Préstamo por Nivel Educativo



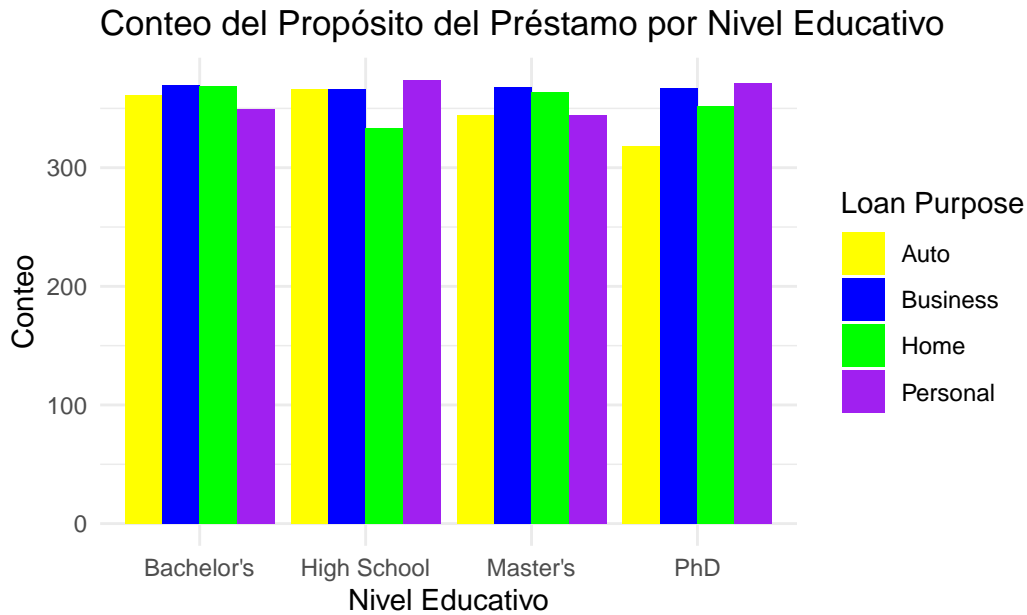
Fuente: Elaboración propia utilizando la base de datos de Kaggle

Por otro lado, ahora queremos ver el gráfico de la variable propósito del préstamo en relación con el grado académico, al ser ambas variables categóricas, lo recomendado es utilizar una geometría que se adapte a esto, Sin embargo, al ser medidas que están muy cercanas, casi no se aprecia la diferencia, por lo que se decide adaptarlo a un gráfico de barras y apreciar mejor la diferencias de manera visual, también recurrimos al uso de colores, para identificar las variables.

```
library(ggplot2)
library(dplyr)

# Contar las combinaciones de Education Level y Loan Purpose
Base_count <- Base_limpia %>%
  group_by(`Education Level`, `Loan Purpose`) %>%
  summarise(Count = n(), .groups = 'drop')

# Gráfico de barras para visualizar el conteo
ggplot(Base_count, aes(x = `Education Level`, y = Count, fill = `Loan Purpose`)) +
  geom_bar(stat = "identity", position = "dodge") + # Usa barras para mostrar el conteo
  labs(x = "Nivel Educativo", y = "Conteo", title = "Conteo del Propósito del Préstamo por N")
  scale_fill_manual(values = c("Business" = "blue", "Personal" = "purple", "Home" = "green",
  theme_minimal() +
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

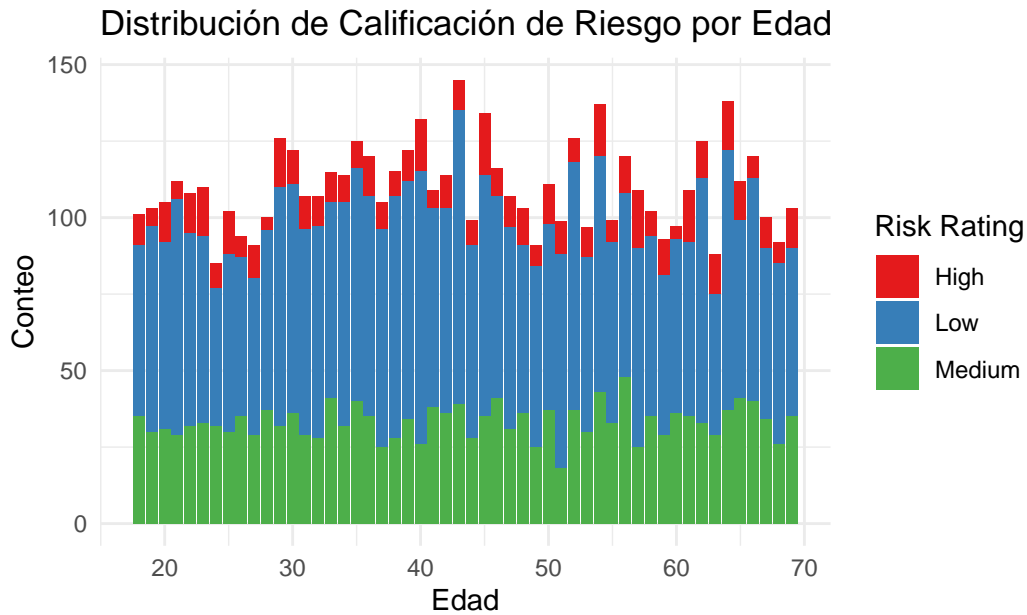


Fuente: Elaboración propia utilizando la base de datos de Kaggle

Para terminar esta sección, vamos a graficar con la variable objetivo de interés, la cual es la Calificación de riesgo, para ello, lo vamos a comparar a través de diversos gráficos con las variables de nivel educativo, ingresos, monto del préstamo, edad y género. Es importante mencionar, que las variables en conjunto afectan a esta calificación, al menos así es de manera teórica. En este apartado nos vamos a centrar en estos gráficos, en secciones posteriores nos encargaremos de hacer la conexión entre la teoría, nuestras hipótesis y los datos obtenidos.

Para comenzar, nos interesa observar cómo se comporta la variable de Calificación de Riesgo, con respecto a la edad:

```
library(ggplot2)
ggplot(Base_limpia, aes(x = Age, fill = `Risk Rating`)) +
  geom_bar() +
  labs(x = "Edad", y = "Conteo", title = "Distribución de Calificación de Riesgo por Edad") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set1") +
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

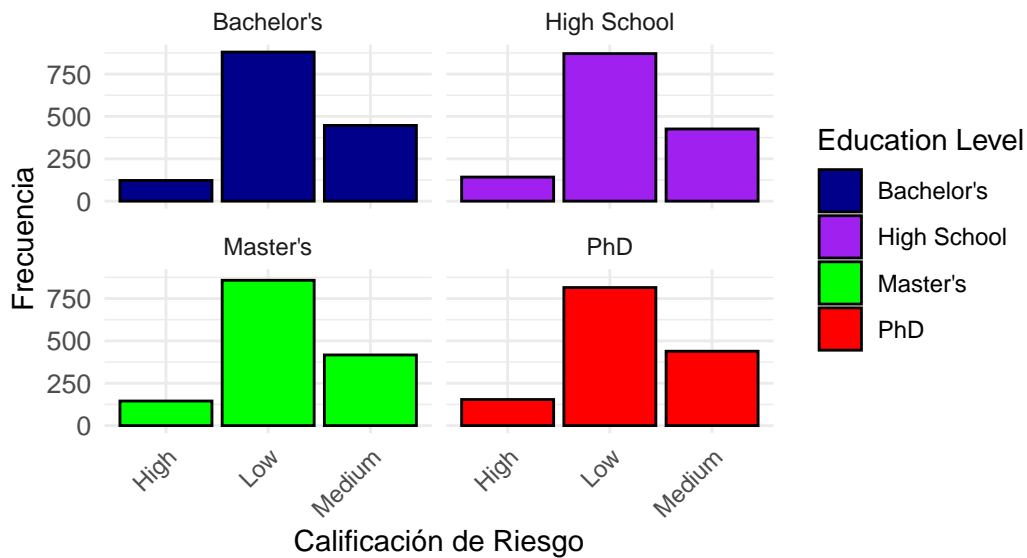


Por otro lado, vamos a ver cómo se comporta la distribución de la calificación de riesgo, haciendo un facet con la categoría Nivel Educativo. Al ser dos variables categóricas las que estamos comparando, lo recomendado es utilizar un gráfico de barras.

```
library(ggplot2)

ggplot(Base_limpia, aes(x = `Risk Rating`, fill = `Education Level`)) +
  geom_bar(color = "black", position = "identity") +
  scale_fill_manual(values = c("Bachelor's" = "darkblue", "High School" = "purple", "Master's" = "green")) +
  facet_wrap(~ `Education Level`) +
  labs(x = "Calificación de Riesgo", y = "Frecuencia", title = "Distribución de la Calificación de Riesgo por Nivel Educativo") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10), # Ajustamos el tamaño del texto
        axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotamos el texto
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

## Distribución de la Calificación de Riesgo por Nivel Educativo



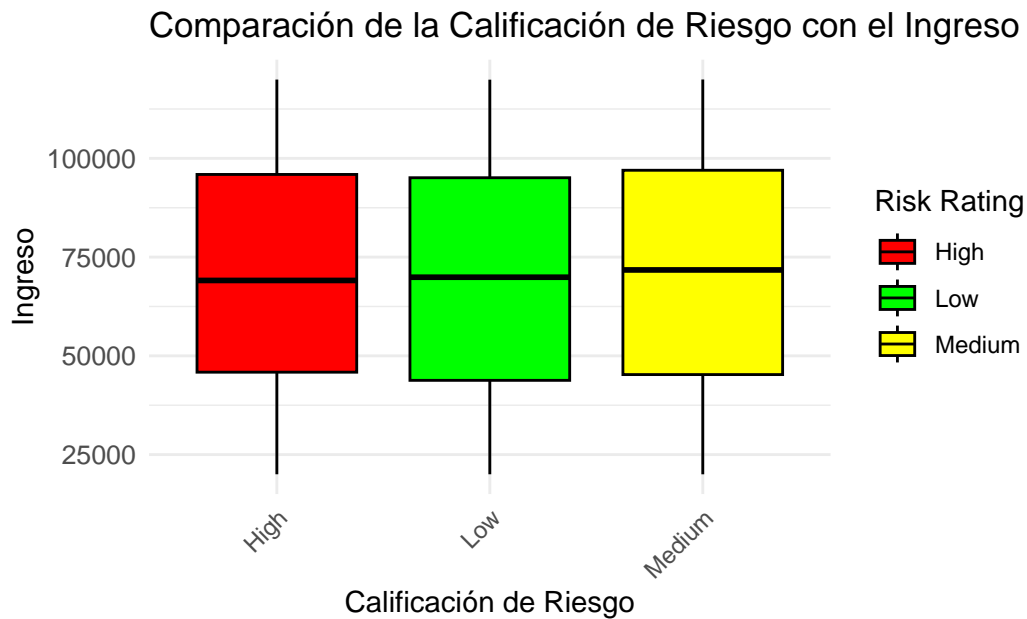
Fuente: Elaboración propia utilizando la base de datos de Kaggle

A su vez, nos interesa observar cómo se ve el gráfico de la variable calificación de riesgo con la variable contra la variable ingreso. Para este gráfico, vamos a utilizar el recomendado en las notas del profesor Maikol Solís, el cual indica que usar diagramas de cajas es útil para comparar las distribuciones.

```
library(ggplot2)

ggplot(Base_limpia, aes(x = `Risk Rating`, y = Income, fill = `Risk Rating`)) +
  geom_boxplot(color = "black") +
  scale_fill_manual(values = c("Low" = "green", "Medium" = "yellow", "High" = "red", "Very High" = "blue")) +
  labs(x = "Calificación de Riesgo", y = "Ingreso", title = "Comparación de la Calificación de Riesgo por Nivel Educativo") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10), # Ajustamos el tamaño del texto
        axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotamos el texto
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```



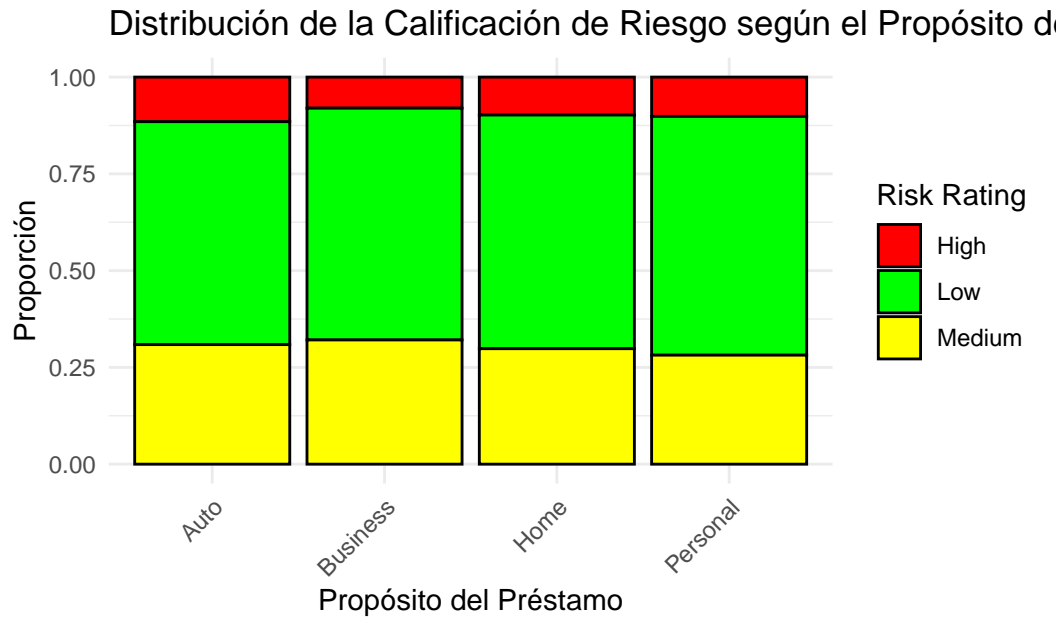


Fuente: Elaboración propia utilizando la base de datos de Kaggle

Otra de las variables de interés, es la calificación de riesgo contra el propósito del préstamo.

```
library(ggplot2)

ggplot(Base_limpia, aes(x = `Loan Purpose`, fill = `Risk Rating`)) +
  geom_bar(position = "fill", color = "black") +
  scale_fill_manual(values = c("Low" = "green", "Medium" = "yellow", "High" = "red", "Very High" = "red")) +
  labs(x = "Propósito del Préstamo", y = "Proporción", title = "Distribución de la Calificación de Riesgo") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotamos el texto para mejorar la legibilidad
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```

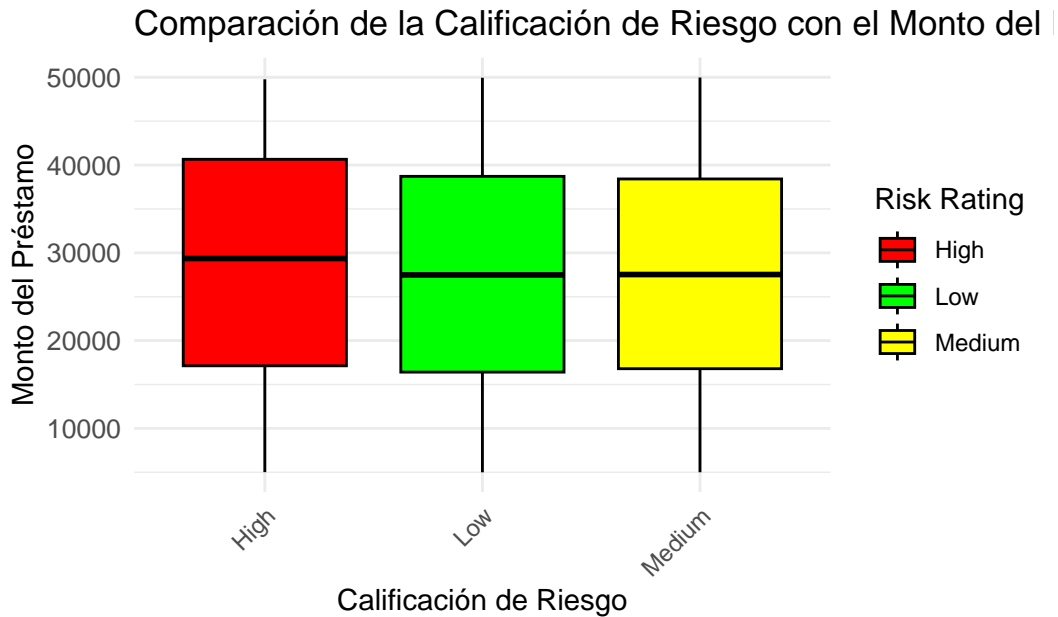


Fuente: Elaboración propia utilizando la base de datos de Kaggle

Por último, vamos a ver la distribución que existe de la variable calificación de riesgo cuando la comparamos contra el monto del préstamo.

```
library(ggplot2)

ggplot(Base_limpia, aes(x = `Risk Rating`, y = `Loan Amount`, fill = `Risk Rating`)) +
  geom_boxplot(color = "black") +
  scale_fill_manual(values = c("Low" = "green", "Medium" = "yellow", "High" = "red", "Very High" = "blue")) +
  labs(x = "Calificación de Riesgo", y = "Monto del Préstamo", title = "Comparación de la Calificación de Riesgo contra el Monto del Préstamo") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10), # Ajustamos el tamaño del texto
        axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotamos el texto
  labs(caption = "Fuente: Elaboración propia utilizando la base de datos de Kaggle") +
  theme(plot.caption = element_text(hjust = 0.5))
```



## 2.4 Propuesta Metodológica

se utilizan los gráficos, como ayuda en las distribuciones de las variables, con el fin de observar su comportamiento. Sin embargo, hay que hacer un análisis más a profundidad, para ello utilizamos el coeficiente de correlación de Pearson, que fue justo el que utilizamos para crear nuestra matriz de correlación, a continuación enunciamos el proceso teórico de dicho procedimiento.

Como menciona el autor (Edgar Apaza 2022) “Los análisis de correlación son métodos estadísticos descriptivos utilizados en investigación de nivel relacional, con los que estima la magnitud y define la tendencia de la relación entre variables.”. Como queremos encontrar alguna relación en nuestra variables de interés, entonces queremos utilizar esta metodología para encontrar dicha relación. Según este mismo autor “el método de correlación de Pearson es una técnica bivariada que emplea en circunstancia multivariada para la explicación de diversos fenómenos relacionados en el campo animal y vegetal. En la correlación de Pearson, los procedimientos guardan relación con la naturaleza de las variables utilizadas.”. Como podemos observar el método nos sirve para este estudio, pues tenemos muchas variables, pero el índice sale de comparar dos a dos las variables, para obtener la correlación entre ellas, lo que luego acomodamos en una matriz para tener una mejor observación de ellas.

R ya posee una librería que calcula automáticamente este coeficiente, sin embargo, la manera teórica de hacerlo es mediante la fórmula:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

En donde  $\rho$  recibe el nombre de Coeficiente de correlación de Pearson y además tiene que ocurrir que  $-1 \leq \rho \leq 1$ . Según el autor (Edgar Apaza 2022) "... se deduce la magnitud de la relación lineal entre variables, los que pueden ser calificados como: Correlación Nula (0), Muy baja (0.01 a 0.1), Débil (0.11 a 0.5), Media (0.51 a 0.75), considerable (0.76 a 0.9), Muy fuerte (0.91 a 0.99) y Perfecta(1)". Otro aspecto importante a decir acerca de nuestro estudio y que tiene relación con lo que este mismo autor afirma y es que " $\rho = 0.000$ , no necesariamente implica que no exista relación entre las variables, sino que la relación podría ser no lineal". Esto de hecho es un buen punto en vista de los resultados que arrojó nuestra matriz de correlación. Además, algo que deberíamos de tomar en cuenta, es que estamos trabajando con variables categóricas, a la hora de la conversión puede haber fuga de información, por ello hay que tener cuidado con solo ver una cifra y animarse a dar una conclusión, cuando en realidad hay que analizar a detalle qué es lo que está pasando.

## 2.5 Construcción de Fichas de Resultados

```
if (!requireNamespace("kableExtra", quietly = TRUE)) {
  install.packages("kableExtra")
}
library(kableExtra)

data <- data.frame(
  Encabezado = c("Nombre de Su hallazgo",
                 "Resumen en una Oración",
                 "Problemas o Posibles Desafíos",
                 "Resumen en un párrafo"),
  Contenido = c("Poca Correlación entre las variables.",
                "Encontramos que las variables a utilizar no presentan una correlación fuerte",
                "La conversión de variable categórica a variable numérica puede estar afectando",
                "Al utilizar el índice de correlación de Pearson, hay que utilizar variables")
)

if (knitr::is_html_output()) {
  # Si es HTML
  kable(data, col.names = c("Encabezado", "Contenido"),
        format = "html",
        escape = FALSE) %>%
  kable_styling(full_width = FALSE) %>%
```

```

    add_header_above(c("Hallazgo de Resultado 1" = 2), bold = TRUE)
} else {
  # Si es PDF
  kable(data, col.names = c("Encabezado", "Contenido"),
        format = "latex",
        booktabs = TRUE) %>%
    kable_styling(latex_options = c("striped", "hold_position"))
}

```

Encabezado	Contenido
Nombre de Su hallazgo	Poca Correlación entre las variables.
Resumen en una Oración	Encontramos que las variables a utilizar no presentan una correlación fuerte,
Problemas o Posibles Desafíos	La conversión de variable categórica a variable numérica puede estar afectando
Resumen en un párrafo	Al utilizar el índice de correlación de Pearson, hay que utilizar variables numéricas

```

if (!requireNamespace("kableExtra", quietly = TRUE)) {
  install.packages("kableExtra")
}
library(kableExtra)

data <- data.frame(
  Encabezado = c("Nombre de Su hallazgo",
                 "Resumen en una Oración",
                 "Problemas o Posibles Desafíos",
                 "Resumen en un párrafo"),
  Contenido = c("Datos muy Iguales",
               "A la hora de graficar las variables se puede observar que los datos tienen poca correlación",
               "Este problema se puede estar ocasionando debido a que las variables de la base de datos no son numéricas",
               "Al realizar las diferentes gráficas, con el objetivo de observar la distribución de los datos")
)

if (knitr::is_html_output()) {
  # Si es HTML
  kable(data, col.names = c("Encabezado", "Contenido"),
        format = "html",
        escape = FALSE) %>%
    kable_styling(full_width = FALSE) %>%
    add_header_above(c("Hallazgo de Resultado 2" = 2), bold = TRUE)
} else {

```

```
# Si es PDF
kable(data, col.names = c("Encabezado", "Contenido"),
      format = "latex",
      booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "hold_position")) %>%
  add_header_above(c("Hallazgo de Resultado 2" = 2), bold = TRUE)
}
```

## Hallazgo de Resultado 2

Encabezado	Contenido
Nombre de Su hallazgo	Datos muy Iguales
Resumen en una Oración	A la hora de graficar las variables se puede observar que los datos tienen dist
Problemas o Posibles Desafíos	Este problema se puede estar ocasionando debido a que las variables de la ba
Resumen en un párrafo	Al realizar las diferentes gráficas, con el objetivo de observar la distribución o

Para la siguiente ficha, es de importancia hacer la acotación, que este curso evalúa el tratamiento de datos, más que el tema de la investigación, pues es un curso de herramientas de datos, por ello, se nos hace pertinente mencionar que el tratamiento de las variables categóricas con variables numéricas puede llevar a problemas, sino se hace un buen tratamiento, además como mencionamos anteriormente en la metodología, que las variables presenten poca correlación, se puede deber a que la relación entre ellas no es lineal o al menos no es cuantificable, pero a la hora de trabajar o de interpretarlas, si tiene sentido hacerlo o existe un cuerpo teórico que lo apoya. A continuación el cuadro del hallazgo:

```
if (!requireNamespace("kableExtra", quietly = TRUE)) {
  install.packages("kableExtra")
}
library(kableExtra)

data <- data.frame(
  Encabezado = c("Nombre de Su hallazgo",
                 "Resumen en una Oración",
                 "Problemas o Posibles Desafíos",
                 "Resumen en un párrafo"),
  Contenido = c("Dificultad a la hora de trabajar con variables categóricas y numéricas.",
                "Buscar correlaciones en variables que no son del mismo tipo puede ocasionar",
                "A la hora de tratar variables categóricas con variables numéricas, se puede",
                "Durante el proceso de tratamiento de datos, hemos observado como los datos s
  )
```

```

if (knitr::is_html_output()) {
  # Si es HTML
  kable(data, col.names = c("Encabezado", "Contenido"),
        format = "html",
        escape = FALSE) %>%
    kable_styling(full_width = FALSE) %>%
    add_header_above(c("Hallazgo de Resultado 3" = 2), bold = TRUE)
} else {
  # Si es PDF
  kable(data, col.names = c("Encabezado", "Contenido"),
        format = "latex",
        booktabs = TRUE) %>%
    kable_styling(latex_options = c("striped", "hold_position")) %>%
    add_header_above(c("Hallazgo de Resultado 3" = 2), bold = TRUE)
}

```

Hallazgo de Resultado 3	
Encabezado	Contenido
Nombre de Su hallazgo	Dificultad a la hora de trabajar con variables categóricas y numéricas.
Resumen en una Oración	Buscar correlaciones en variables que no son del mismo tipo puede ocasionar
Problemas o Posibles Desafíos	A la hora de tratar variables categóricas con variables numéricas, se puede es
Resumen en un párrafo	Durante el proceso de tratamiento de datos, hemos observado como los datos

## **3 Bitácora 3**

### **3.1 Parte de Planificación**

#### **3.1.1 Análisis de modelación**

#### **3.1.2 Construcción de fichas de resultados**

#### **3.1.3 Ordenamiento de los elementos de reporte**

### **3.2 Parte de escritura**

### **3.3 Parte de reflexión**



## 4 Anexo

### 4.1 Anexo 1 (CHANGELOG Bitacora 1)

#### 4.1.1 Chore

- Agrega archivo configuracion pre-commit
- Agrega configuracion repo-actions
- Agrega cambios en docs/
- Modifica la carpeta docs
- Modifica la carpeta docs
- Modifica la carpeta docs

#### 4.1.2 Feat

- Agrega documentos para cuarto
- Agrega carpeta docs
- Agrega la base de datos en formato csv
- Agrego la base de datos en formato txt a manera de respaldo del archivo en formato csv. Cualquier cambio realizado al archivo original sera realizado en este tambien.
- Agrega documentos Bitacora\_1
- Agrega comentario en Bitacora\_1
- Elimina la base de datos
- Agrega la base de datos
- Agrega la definición de la idea
- Agrega en bitácora 1 identificación de tensiones
- Agrega en bitácora 1, identificación de Tensiones
- Agrega en bitácora 1, Reformulación de la idea en modo pregunta
- Agrega en bitácora 1, argumentación de las preguntas
- Agrega en bitácora 1, argumentación a través de datos
- Agrega en Bitacora 1 avance de revisión bibliografica
- Agrega en bitácora 1, parte de escritura
- Agrega el archivo de references.bib
- Agrega las referencias bibliograficas
- Agrega en bitacora 1 principios y teorias

- Agrega en bitacora 1 busqueda bibliografica
- Agrega la introduccion
- Se agrega a bitácora 1, la UVE de Gowin
- Agrega imagen de la V de Gowin

### 4.1.3 Fix

- Arregla una funcionalidad de .gitignore
- Corrige la numeración de la bitacora 1
- Arreglo de index
- Realiza correcciones diversas en los documentos
- Corrige error ortografico

## 4.2 Anexo 2 (Participacion Bitacora 1)

```

project      : Bitacora-Grupo-5-CA-0204-II-2024-
repo age     : 7 days
branch:      : main
last active  : 40 minutes ago
active on    : 4 days
commits      : 36
files        : 41
uncommitted  : 10
authors      :
    26 Jeikel Navarro 72.2%
    8  Erick Venegas  22.2%
    2  Gara           5.6%

```

Figura 4.1: Summary Bitacora 1

## 4.3 Anexo 3 (CHANGELOG Bitacora 2)

```
### Chore

- Agrega archivo configuracion pre-commit
- Agrega configuracion repo-actions
- Agrega cambios en docs/
- Modifica la carpeta docs
- Modifica la carpeta docs
- Modifica la carpeta docs
- Modifica cuarto para agregar bitacora 2
```

Figura 4.2: Chore Bitacora 2

```
### Feat

- Agrega documentos para cuarto
- Agrega carpeta docs
- Agrego la base de datos en formato csv
- Agrego la base de datos en formato txt a manera de respaldo del archivo en formato csv. Cualquier cambio realizado al archivo original sera realizado en este tambien.
- Agrega documentos Bitacora_1
- Agrega comentario en Bitacora_1
- Elimina la base de datos
- Agrega la base de datos
- Agrega la definición de la idea
- Agrega en bitacora 1 identificación de tensiones
- Agrega en bitacora 1, identificación de Tensiones
- Agrega en bitacora 1, Reformulación de la idea en modo pregunta
- Agrega en bitacora 1, argumentación de las preguntas
- Agrega en bitacora 1, argumentación a través de datos
- Agrega en Bitacora 1 avance de revisión bibliografica
- Agrega en bitacora 1, parte de escritura
- Agrega el archivo de referencias.bib
- Agrega las referencias bibliograficas
- Agrega en bitacora 1 principios y teorías
- Agrega en bitacora 1 busqueda bibliografica
- Agrega la introducción
- Se agrega a bitacora 1, la UVE de Gowin
- Agrega imagen de la V de Gowin
- Agrega archivos necesarios para el changelog y summary
- Agrega titulos y contenido en el análisis estadístico
- Agrega base de datos
- Agrega datos resumidos de las variables
- Agrega nuevas referencias
- Agrega variables resumen de interés
- Agrega gráfico de distribución por género
- Agrega gráfica de la variable ingreso
- Agrega gráfico de la variable ingreso y facet por nivel académico
- Agrega gráfico de la variable ingreso y facet por género
- Agrega gráfico de la variable monto del préstamo facet por nivel educativo y género
- Agrega gráfico de la variable propósito del préstamos por nivel educativo
- Agrega gráfico de la variable calificación por riesgo contra edad
- Se agrega el archivo html de la bitacora 1
- Se actualiza el archivo .css de la bitacora 1
- Se agrega una nueva ficha de literatura
- Se agrega una ficha de literatura
- Se actualiza el archivo html de la bitacora 1
- Agrega gráfico de variable calificación de riesgo facet nivel educativo
- Agrega gráfico de cajas de variable calificación de riesgo contra salario
- Agrega gráfico de la variable calificación de riesgo contra propósito del préstamo
- Agrega la tabla de organización y literatura
- Agrega el archivo de Apéndice
- Se agrega el archivo de anexo
- Se agregan subtítulo y nuevos capítulos
- Se agregan nuevas referencias
- Se agrega una nueva ficha de literatura
- Se actualiza el archivo html de la bitacora 1
- Se agrega el primer enlace de literatura
- Se agregan 2 enlaces de literatura
- Se agregan los últimos enlaces de literatura
```

Figura 4.3: Feat Bitacora 2

```
### Fix
- Arregla una funcionalidad de .gitignore
- Corrige la numeración de la bitacora 1
- Arreglo de index
- Realiza correcciones diversas en los documentos
- Corrige error ortografico
- Correccion de error en el resumen.txt
- Se corrige la definicion de la idea
```

Figura 4.4: Fix Bitacora 2

## 4.4 Anexo 4 (Participacion Bitacora 2)

```
$ git summary

project      : Bitacora-Grupo-5-CA-0204-II-2024-
repo age     : 5 weeks
branch:      : main
last active  : 10 minutes ago
active on    : 8 days
commits      : 72
files        : 50
uncommitted  : 2
authors      :
    30 Jeikel Navarro  41.7%
    23 Erick Venegas   31.9%
    19 Gara            26.4%
```

Figura 4.5: Summary Bitacora 2

## 5 Apendice

### 5.1 Correcciones Bitácora 1

La correccion más relevantes se detalla a continuación:

- **Corrección 1:** Se ha modificado la sección de definición de la idea para ajustarnos mejor a lo que se espera ser observado en la misma. Tal que, nos centramos más en definir claramente la idea de investigación, sin entrar en los detalles ni en dar el paso a paso de cómo se lleva la misma acabo. Por lo que, en esta sección, solo nos limitamos a exponer qué es lo que buscamos investigar y por qué es importante o relevante, dejando los aspectos metodológicos y el desarrollo de este para otras secciones.

### 5.2 Sugerencias Bitácora 1

A partir del foro consideramos todas las sugerencias:

1. *El grupo podría ser más explícito en la metodología que usará y como abordará el objetivo general, esto lo observé en la sección de principios y teorías donde me parece que no tienen muy claro una teoría con la cual construir ideas relacionadas a la investigación. Hacen un buen análisis de la relación del tema de Riesgos con la Probabilidad y algunas teorías de valoración de riesgos financieros, pero podrían ser más explícitos en cual usarán y como la usarán.*
- **Respuesta:** Gracias por la observación. Trataremos de que la metodología a utilizar quede un poco más explícita en la sección de principios y teorías, especialmente en relación con el análisis de la valoración de riesgos.
2. *Es crucial asegurar la validez de los datos utilizados, ya que estos deben reflejar con precisión las características reales de las personas con calificación crediticia. Aunque el enfoque del trabajo esté en el uso de herramientas de análisis de datos, es fundamental que la base de datos sea coherente con la realidad, dado que las conclusiones dependerán de la calidad de la información analizada. Además, sería recomendable incluir un título claro y descriptivo que refleje el objetivo principal del análisis.*

- **Respuesta:** Agradecemos la sugerencia. Creemos que la base de datos que hemos escogido tiene una gran calidad o por lo menos visto desde un punto de vista académico, ya que está respaldada por el Instituto Tecnológico de Massachusetts (MIT). Adicional a eso, también hemos hecho una limpieza de la base de datos para tratar de asegurarnos de aquellos datos tengan cierto grado de validez. A su vez, también hemos dejado mas en claro el título de la investigación.
3. *Considero que el proceso de selección de la pregunta final se pudo haber beneficiado de una argumentación más detallada para cada una de las opciones preliminares, especialmente haciendo una distinción más explícita entre los factores lógicos, éticos y emocionales que respaldan o cuestionan cada posible interrogante. Asimismo, es posible que la parte final de esta bitácora se pueda enriquecer con una mayor diversidad de fuentes, lo que ofrecería una perspectiva más amplia sobre los estudios que se han realizado previamente sobre el tema.*
- **Respuesta:** Se tomará en cuenta su comentario. A la vez, estamos de acuerdo en que tal vez no llegamos a abarcar una gran variedad de fuentes, pero esto es debido a que la investigación se quiere mantener lo más aterrizada posible, de manera que no haya fuentes que sean innecesarias que a lo largo del trabajo no lleguen a tener relevancia. No obstante, hemos decidido ampliar un poco más las fuentes, manteniendo el principio de no meter fuentes innecesarias.
4. *Sentí que la bitácora les quedó muy corta, la leí en poco tiempo y no entendí muy bien a qué querían llegar con los objetivos planteados. También siento que les faltó un poco más de referencias. La ventaja es que gracias a la naturaleza del tema esto se puede mejorar bastante fácil en las siguientes bitácoras.*
- **Respuesta:** Agradecemos sus comentarios. A lo largo de la segunda bitácora vamos a profundizar más en los objetivos planteados, además, talvez la extensión de la primer bitácora se optó que fuese lo más concisa posible, tratando de que quedase lo más claro posible lo planteado en la misma, talque no se llegara a caer en redundancias. También como se menciona en otro comentario añadiremos más referencias que sean de utilidad para respaldar mejor el análisis.

### 5.3 Referencias bibliográfica

- Chicu, Dorina. 2020. «La valoración del riesgo financiero». 2020. <https://openaccess.uoc.edu/bitstream/10609/150126/1/LaValoracionDelRiesgoFinanciero.pdf>.
- Edgar Apaza, César Condori, Samuel Cazorla. 2022. «La Correlación de Pearson o de Spearman en caracteres físicos y textiles de la fibra de alpacas». 2022. <http://www.scielo.org.pe/pdf/rivep/v33n3/1609-9117-rivep-33-03-e22908.pdf>.

- Hadley Wickham, Garrett Golemund. 2019. «R for Data Science (2nd ed.)». 2019. <https://digitallibrary.tsu.ge/book/2019/september/books/R-for-Data-Science.pdf>.
- Maria de los Ángeles Herrera, Juan Terán. 2024. «Conceptualización del riesgo de los mercados financieros». 2024. <https://www.redalyc.org/pdf/900/90075920006.pdf>.
- Palacios, Alberto. 2012. «Calificación de riesgo: definición e influencia en la última década». 2012. <https://digibuo.uniovi.es/dspace/bitstream/handle/10651/4017/ACC-.pdf;jsessionid=723581A47435AFB6D2FEC05A70379F77?sequence=1>.
- Solis, Maicol. 2024. «Guía del curso: Herramienta para Ciencia de Datos». 2024. <https://maikolsolis.com/libros/hpcd/>.