

Índice

Introducción	3
Objetivos	4
Objetivo N°1	4
Objetivo N°2	4
Objetivo N°3	4
Base de Datos	4
Descripción de las variables de la base de datos	4
Monto del préstamo (“loan_amount”)	4
Tiempo de empleo (emp_length)	5
Propiedad de vivienda (home_ownership)	5
Nivel de ingresos(income)	5
Propósito del préstamo (loan_purpose)	5
Nivel de endeudamiento(debt_to_income)	5
Puntaje crediticio (credit_score)	5
Consulta reciente (recent_inquiry)	5
Registro de morosidad (delinquent)	5
Cuentas de crédito (credit_accounts)	6
Antecedentes públicos negativos (bad_public_record)	6
Utilización de crédito (credit_utilization).	6
Bancarrota previa (past_bankrupt)	6
Análisis de Resultados	6
Análisis de Resultados del Objetivo N°1	6
Análisis Univariante	6
Análisis Descriptivo	7
Análisis Inferencial: Comparación de la proporción de nivel de endeudamiento alto entre propietarios e inquilinos	7
Conclusiones	8
Análisis de Resultados del Objetivo N°2	9
Análisis univariante	9
Análisis Descriptivo	10
Análisis Inferencial: Comparación de la proporción de personas con un puntaje crediticio alto que se han declarado en quiebra.	11
Conclusiones	11
Análisis de Resultados del Objetivo N°3	11

Analisis Univariante	12
Análisis Descriptivo	13
Análisis Inferencial: Contrastar la proporción del monto de préstamo de nivel alto entre los prestatarios que registran un record público negativo y aquellos que no.	17
Conclusiones	19
Referencias	20
Anexos	21

Introducción

En un mundo en constante evolución financiera, los préstamos bancarios se han convertido en una herramienta esencial para que individuos y empresas alcancen sus objetivos económicos. Los préstamos no solo tienen un impacto directo en la vida financiera de las personas, sino que también juegan un papel fundamental en la economía en general. Analizar y comprender los datos relacionados con préstamos bancarios puede brindar valiosas perspectivas sobre patrones de endeudamiento, riesgos financieros y comportamientos económicos.

La aprobación de un préstamo depende de una serie de factores que los bancos evalúan meticulosamente para minimizar riesgos y asegurar la viabilidad de la inversión. Según José Luis Martínez, portavoz de la Asociación Española de Banca (AEB), el éxito de la operación a largo plazo. “La prioridad de los bancos es conceder préstamos de forma responsable y transparente, con el doble objetivo final de dar el mejor servicio al cliente y también protegerlo”

El presente proyecto tiene como objetivo realizar un análisis estadístico inferencial exhaustivo sobre un conjunto de datos de préstamos bancarios. A través de la aplicación de técnicas de estadística inferencial, buscaremos no solo describir las características y tendencias presentes en los datos, sino también extraer conclusiones que permitan hacer inferencias significativas sobre la población de préstamos en cuestión.

Objetivos

Objetivo N°1

Comparar la proporción de nivel de endeudamiento alto entre prestatarios propietarios e inquilinos.

Objetivo N°2

Comparar la proporción del nivel alto de puntuaciones de crédito entre aquellos prestatarios que se han declarado en quiebra y los que no.

Objetivo N°3

Contrastar la proporción del monto de préstamo de nivel alto entre los prestatarios que registran un record público negativo y aquellos que no.

Base de Datos

Monto del Prestamo	Tiempo de empleo	Propiedad de la vivienda	Nivel de ingresos	Propósito del préstamo	Nivel de endeudamiento	Puntaje de crédito	Registro de Consultas recientes	Cuentas de crédito morosas	Registro público de crédito negativo	Utilización de crédito	Pasado en bancarrota
LOW	10+ years	RENT	LOW	credit_card	HIGH	AVERAGE	YES	NEVER FEW	NO	HIGH	NO
LOW	< 2 years	RENT	LOW	car	LOW	AVERAGE	YES	NEVER FEW	NO	LOW	NO
LOW	10+ years	RENT	LOW	small_business	AVERAGE	AVERAGE	YES	NEVER FEW	NO	HIGH	NO
MEDIUM	10+ years	RENT	MEDIUM	other	HIGH	AVERAGE	YES	MORE THAN 2 YEARS AGO	AVERAGE	LOW	NO
LOW	< 2 years	RENT	HIGH	other	AVERAGE	AVERAGE	NO	MORE THAN 2 YEARS AGO	MANY NO	MEDIUM	NO
LOW	2 - 5 years	RENT	LOW	wedding	AVERAGE	AVERAGE	YES	NEVER	AVERAGE	MEDIUM	NO

Las variables de esta base de datos presentan un valor significativo al evaluar el perfil de los prestatarios y al comprender los factores que pueden afectar su capacidad de reembolso de un préstamo. A continuación, se describe cada variable, junto con las categorías correspondientes que abarcan.

Descripción de las variables de la base de datos

Monto del préstamo (“loan_amount”)

Esta variable cualitativa refleja distintos intervalos o categorías de montos de préstamos que los prestatarios podrían requerir. Estas categorías se dividen en tres: “Bajo” (Low), “Mediano” (Medium) y “Alto” (High),

con el propósito de representar diversos niveles de financiamiento.

Tiempo de empleo (emp_length)

Tiempo de empleo. Esta variable cualitativa indica diferentes categorías de duración de empleo, como “Menos de 1 año”, “1-3 años”, “4-6 años”, etc., para capturar la antigüedad laboral de los prestatarios.

Propiedad de vivienda (home_ownership)

Esta variable cualitativa representa la situación de vivienda de los prestatarios, como “Own”, “Rent” o “Mortgage”, indicando si el prestatario es propietario, inquilino o tiene una hipoteca en su vivienda.

Nivel de ingresos(income)

Esta variable cualitativa se segmenta en tres categorías: “Bajo” (Low), “Mediano” (Medium) y “Alto” (High), con el propósito de representar los distintos niveles de ingresos de los prestatarios.

Propósito del préstamo (loan__ purpose)

Esta variable cualitativa abarca diversas categorías, tales como “Salud” (Medical), “Automóvil” (Car), “Tarjetas de crédito” (Credit Card), “Mejoras para el hogar” (Home Improvement), entre otras. Estas categorías reflejan las distintas razones por las cuales el prestatario está solicitando el préstamo.

Nivel de endeudamiento(debt__to__income)

Esta variable cualitativa tiene el propósito de reflejar los distintos niveles de relación existentes entre la deuda y los ingresos. Los términos utilizados para denotar estos niveles son: “Bajo” (Low), para indicar una proporción baja, “Regular” (Average) para señalar una proporción y “Alto” (High) para representar una proporción elevada. Estos términos capturan la medida en que los ingresos del prestatario están destinados al reembolso de deudas.

Puntaje crediticio (credit_score)

Esta variable cualitativa se encuentra dividida en tres categorías: “Bajo” (Low), Regular (Average) y “Alto” (High), con el propósito de representar distintos niveles de puntuación crediticia.

Consulta reciente (recent__inquiry)

Esta variable cualitativa podría ser “Sí” (Yes) o “No” (No) para indicar si el prestatario ha tenido una consulta reciente en su historial de crédito.

Registro de morosidad (delinquent)

Esta variable cualitativa podría adoptar los siguientes valores: “Nunca” (Never), “Hace más de 2 años” (More than 2 years ago) y “En los últimos 2 años” (In the past 2 years). Estos valores se utilizan para indicar si el prestatario presenta un historial de pagos atrasados.

Cuentas de crédito (credit_accounts)

Esta variable cualitativa representa diferentes niveles de cantidad de cuentas de crédito, como: “Pocas” (Few), “Promedio” (Average), “Muchas” (Many).

Antecedentes públicos negativos (bad_public_record)

Esta variable cualitativa puede adoptar los valores “Sí” (Yes) o “No” (No) con el propósito de señalar si el prestatario posee registros públicos desfavorables en su historial crediticio.

Utilización de crédito (credit_utilization).

Esta variable cualitativa se segmenta en tres categorías: “Bajo” (Low), “Regular” (Average) y “Alto” (High), con el propósito de representar diversos niveles de utilización de crédito en relación a los límites disponibles.

Bancarrota previa (past_bankrupt)

Esta variable cualitativa adquiere los valores “Sí” (Yes) o “No” (No), con el propósito de señalar si el prestatario ha experimentado previamente una bancarrota en su historial crediticio.

Análisis de Resultados

Análisis de Resultados del Objetivo N°1

Objetivo N°1: Comparar la proporción de nivel de endeudamiento alto entre prestatarios propietarios e inquilinos.

Análisis Univariante

Cuadro 2: Tabla de Frecuencias entre prestatarios propietarios e inquilinos con nivel de endeudamiento alto

Var1	Freq
OWN	802
RENT	4966

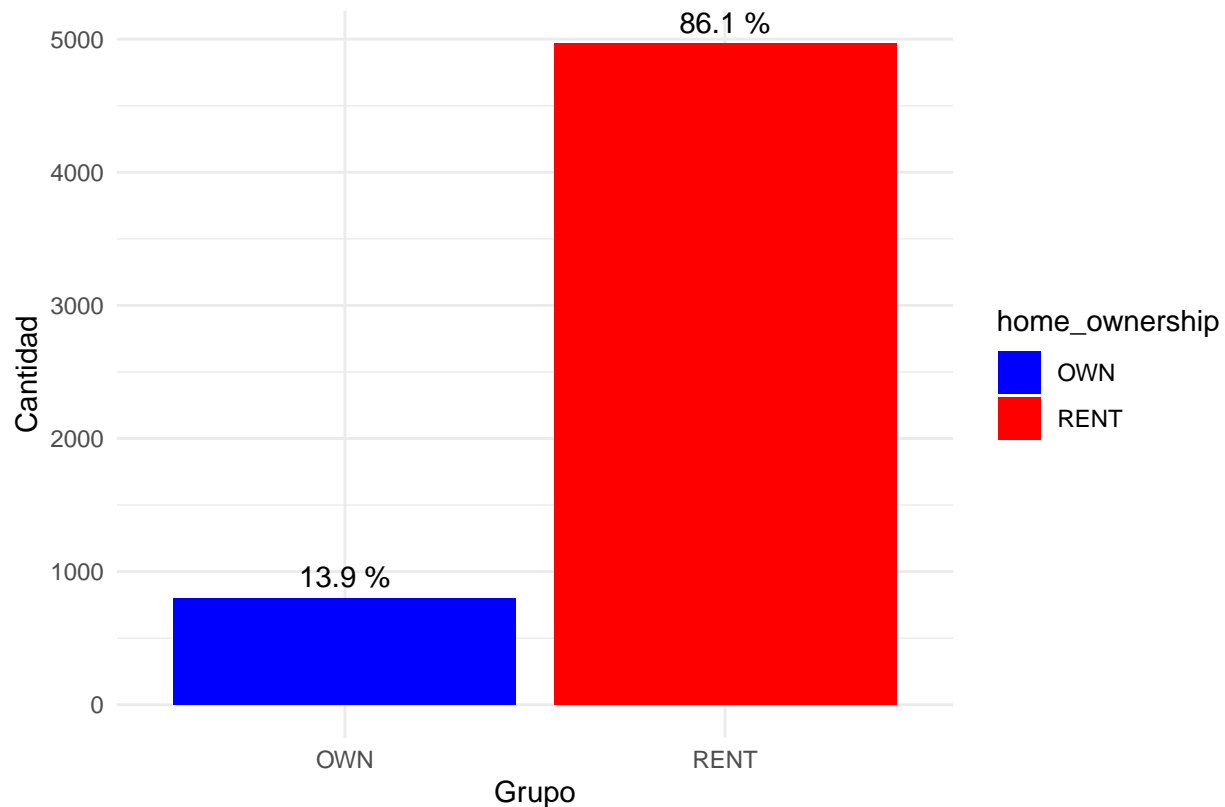


Figura 1. Comparación de Nivel de Endeudamiento Alto entre Propietarios e Inquilinos

Resumen: Al analizar el cuadro 2 y la figura 1, se evidencia que el número de prestatarios que son propietarios asciende a 802, representando el 13.1%. Por ende, el 86.9% restante corresponde a prestatarios que alquilan vivienda, totalizando 4966 individuos, es decir, aproximadamente 6 veces más que aquellos con propiedad propia.

Análisis Descriptivo

El análisis se basa en el cuadro 2 y la figura 1, que proporcionan datos sobre la distribución de prestatarios en términos de propiedad de vivienda. Se ha identificado que un total de 802 prestatarios son propietarios de vivienda, lo que representa el 13.9% del total de prestatarios en el estudio. Esto indica que una minoría relativamente pequeña de prestatarios posee su propia vivienda.

Por otro lado, la gran mayoría, equivalente al 86.1% restante, corresponde a prestatarios que alquilan vivienda. Este grupo está compuesto por 4966 individuos en total. Esto es aproximadamente 6 veces más que la cantidad de prestatarios propietarios. Esta diferencia en la proporción es significativa y sugiere que el alquiler es la forma de tenencia de vivienda más común entre los prestatarios en el estudio.

En resumen, los resultados del análisis muestran una distribución desigual en términos de propiedad de vivienda entre los prestatarios. Un pequeño porcentaje posee su propia vivienda, mientras que la mayoría opta por alquilar. Esto podría reflejar diferentes preferencias y situaciones económicas entre los prestatarios, así como la influencia de factores como el mercado inmobiliario y las condiciones crediticias.

Análisis Inferencial: Comparación de la proporción de nivel de endeudamiento alto entre propietarios e inquilinos

Para realizar la comparación de proporciones, se empleara una prueba de hipótesis para proporciones, siendo:

H_0 : La proporción de inquilinos con nivel de endeudamiento “Alto” es igual a la proporción de propietarios con nivel de endeudamiento “Alto”: $p_1 = p_2$

H_a : La proporción de inquilinos con nivel de endeudamiento “Alto” no es igual a la proporción de propietarios con nivel de endeudamiento “Alto”: $p_1 \neq p_2$

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(nrow(inquilinos_alto_endeudamiento), nrow(proprietarios_alto_endeudamiento)) out of c(total_)
## X-squared = 6009.2, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.7091132 0.7347148
## sample estimates:
##   prop 1   prop 2
## 0.860957 0.139043
```

Usando `prop.test`, el valor de p hallado en la prueba fue igual a $2.2e-16$. En teoría, si este valor es menor a 0.05 se rechaza la hipótesis nula. Por lo tanto, se rechaza que la proporción de inquilinos con nivel de endeudamiento “ALTO” es igual a la proporción de propietarios con nivel de endeudamiento “ALTO”.

Conclusiones

- **Conclusión 1:** Un 13.9% de los prestatarios son propietarios, mientras que un 86.1% son inquilinos. Esta discrepancia sugiere que la mayoría de los clientes del banco opta por el alquiler como forma de vivienda.
- **Conclusión 2:** Los resultados resaltan la tendencia predominante de los clientes del banco hacia el alquiler en lugar de la propiedad de vivienda. Con una proporción significativamente mayor de inquilinos en comparación con propietarios, el banco podría considerar adaptar sus servicios y ofertas para atender a las necesidades específicas de ambos grupos.
- **Conclusión 3:** El análisis de proporciones entre inquilinos y propietarios revela una diferencia estadísticamente significativa en los niveles de endeudamiento “ALTO”. Esto implica que los prestatarios inquilinos y propietarios no presentan las mismas proporciones de endeudamiento “ALTO”, lo que podría tener implicaciones en la gestión de riesgos crediticios por parte del banco.

Análisis de Resultados del Objetivo N°2

Objetivo N°2: Comparar la proporción del nivel alto de puntuaciones de crédito entre aquellos prestatarios que se han declarado en quiebra y los que no.

Análisis univariante

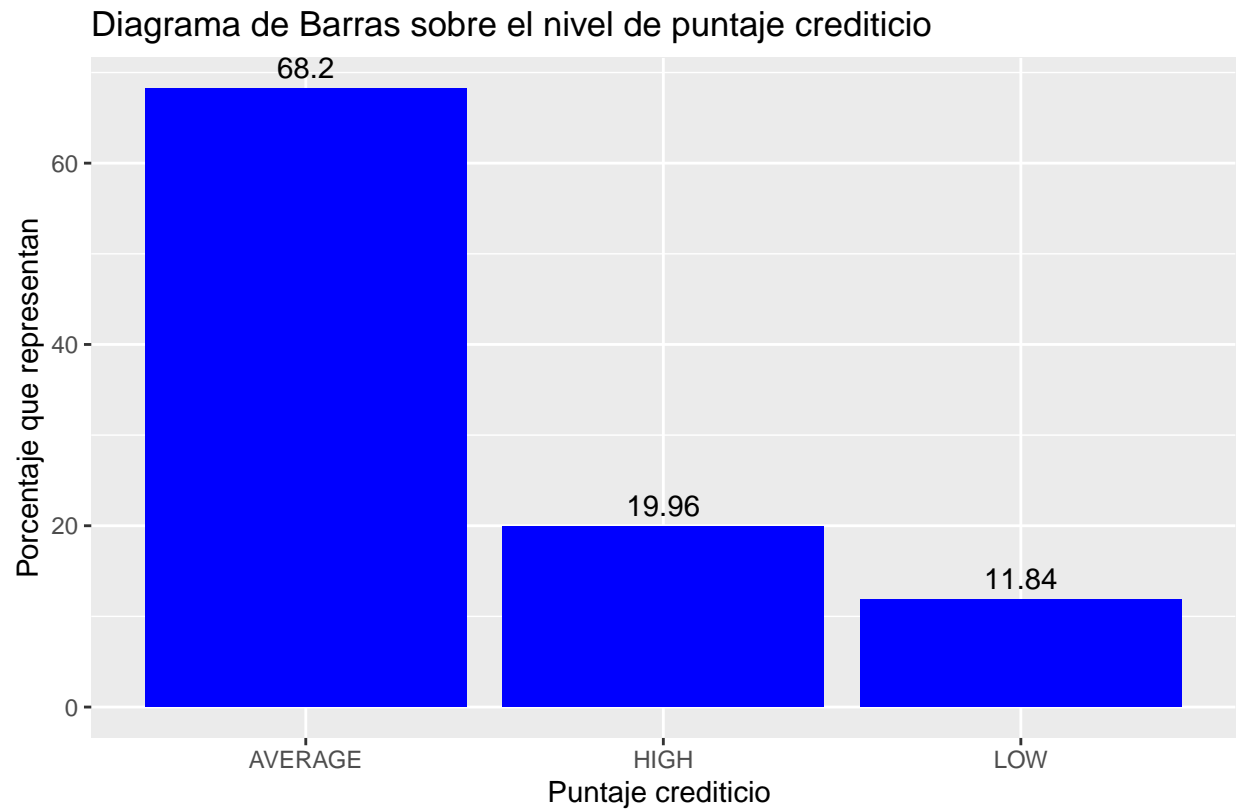


Figura 2. Diagrama de Barras sobre el nivel de puntaje crediticio.

Resumen: En cuanto a la variable “credit_score”, nos damos cuenta que la gran masa de los usuarios tiene un puntaje crediticio medio siendo el 68.2% del total de los usuarios, seguidos por aquellos que tienen un alto puntaje crediticio siendo el 19.96% y la minoría conformada por aquellos con un bajo puntaje crediticio que son el 11.84%

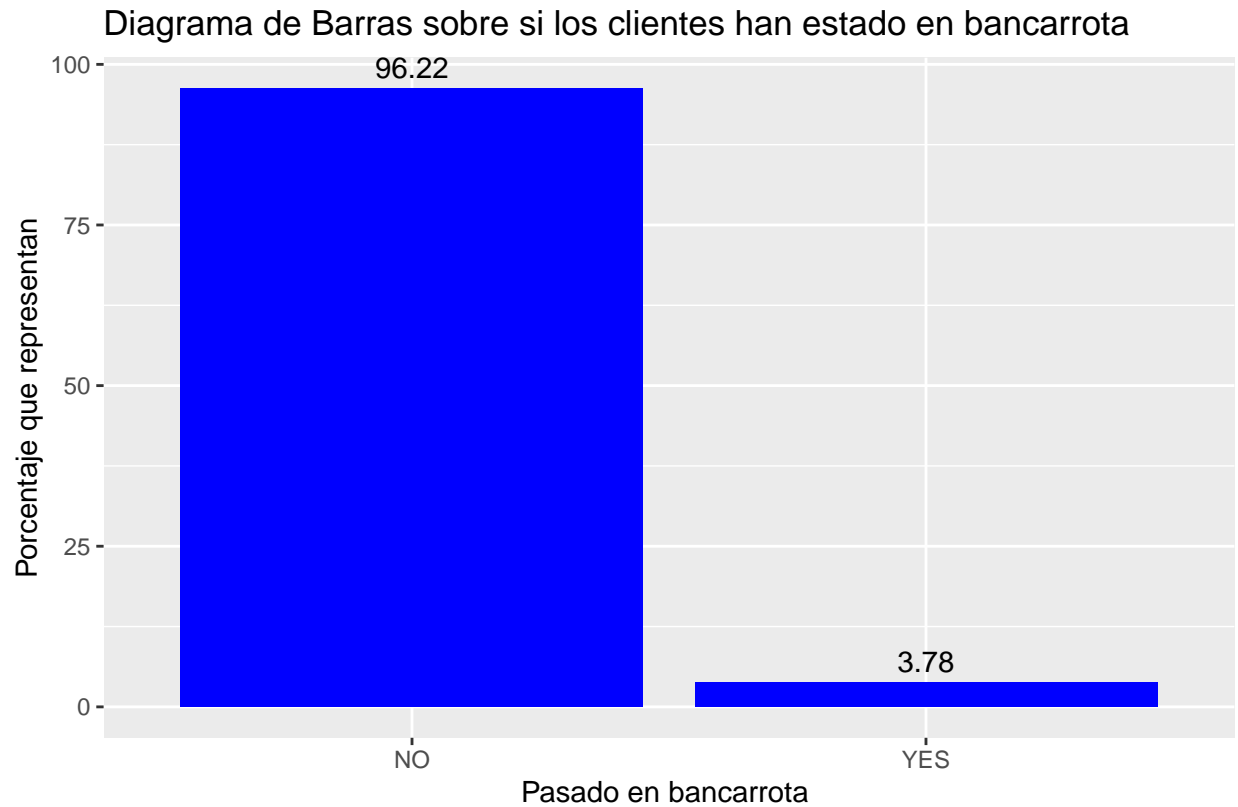


Figura 3. Diagrama de Barras sobre si los clientes han estado en bancarrota

Resumen: En cuanto a la variable “past_bankrupt”, se tiene que el 96.22% de los clientes nunca se han declarado en quiebra, mientras que el 3.78% restante si lo ha hecho , lo que nos da una idea de lo raro que son los usuarios que han estado en quiebra

Análisis Descriptivo

En cuanto a la variable “credit_score”, nos damos cuenta que la gran masa de los usuarios tiene un puntaje crediticio medio siendo el 68.2% del total de los usuarios, seguidos por aquellos que tienen un alto puntaje crediticio siendo el 19.96% y la minoría conformada por aquellos con un bajo puntaje crediticio que son el 11.84% En cuanto a la variable “past_bankrupt”, se tiene que el 96.22% de los clientes nunca se han declarado en quiebra, mientras que el 3.78% restante si lo ha hecho , lo que nos da una idea de lo raro que son los usuarios que han estado en quiebra Dentro de los usuarios con “credit_score=HIGH” o usuarios con un puntaje crediticio alto:

Cuadro 3: Tabla de Frecuencia entre prestatarios de puntaje crediticio alto que han estado en quiebra y aquellos que no

	NO	YES
HIGH	7928	3

Revisando la tabla se puede afirmar que, para la muestra actual de declarados en quiebra con un puntaje crediticio alto, existe una mayor cantidad de quienes no lo hicieron, siendo estos el 99.96% de un total de 7931 personas

Análisis Inferencial: Comparación de la proporción de personas con un puntaje crediticio alto que se han declarado en quiebra.

Para realizar la comparación de proporciones, se empleara una prueba de hipótesis para proporciones, siendo:

H_0 : La proporción de personas con un puntaje crediticio “ALTO” que se han declarado en quiebra es igual a la proporción de personas con un puntaje crediticio “ALTO” que no se han declarado en quiebra: $p_1=p_2$

H_a : La proporción de personas con un puntaje crediticio “ALTO” que se han declarado en quiebra es diferente de la proporción de personas con un puntaje crediticio “ALTO” que no se han declarado en quiebra “Alto”: $p_1 \neq p_2$

```
##
## 1-sample proportions test with continuity correction
##
## data:  tablatotal, null probability 0.5
## X-squared = 7917, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.9989647 1.0000000
## sample estimates:
##           p
## 0.9996217
```

Usando prop.test, el valor de p hallado en la prueba fue igual a 2.2e-16. En teoría, si este valor es menor a 0.05 se rechaza la hipótesis nula. Por lo tanto, se rechaza que la proporción de personas con un puntaje crediticio “ALTO” que se han declarado en quiebra es igual a la proporción de personas con un puntaje crediticio “ALTO” que no se han declarado en quiebra”.

Conclusiones

- **Conclusión 1:** Es muy raro que prestatarios que poseen un alto puntaje crediticio se declaren en quiebra, aun así, estar dentro de esta categoría por increíblemente bajo que sea su porcentaje dentro de la misma no exenta al prestatarios de esa posibilidad, aunque los casos son prácticamente datos aberrantes.
- **Conclusión 2:** Es seguro para la institución financiera ofrecer créditos, ofertas de productos o cualquier otro servicio de este tipo, a prestatarios con un puntaje crediticio alto ya que el porcentaje de estos que quiebran es extremadamente bajo, lo cual prácticamente asegura que la institución financiera recuperara el monto prestado más el interés que le con el que se lo hizo.
- **Conclusión 3:** Se recomienda potenciar la relacion con este tipo de usuarios ya que puede llevar a la generación de referencias de calidad que pueden aumentar la base de clientes ademas de que tener una base de prestatarios con historiales crediticios sólidos puede mejorar la imagen de una entidad financiera en el mercado lo cual aumenta la confianza de los inversores, reguladores y otros interesados en la institución.

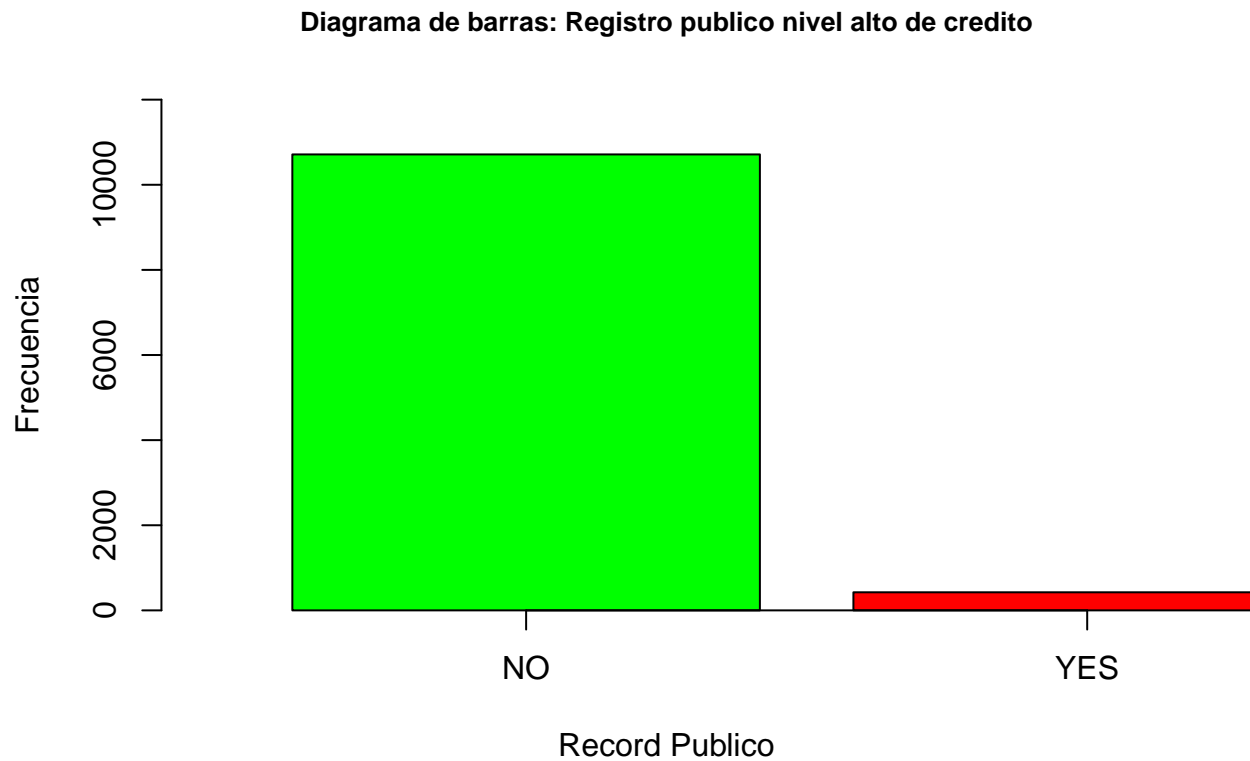
Análisis de Resultados del Objetivo N°3

Objetivo N°3: Contrastar la proporción del monto de préstamo de nivel alto entre los prestatarios que registran un record público negativo y aquellos que no.

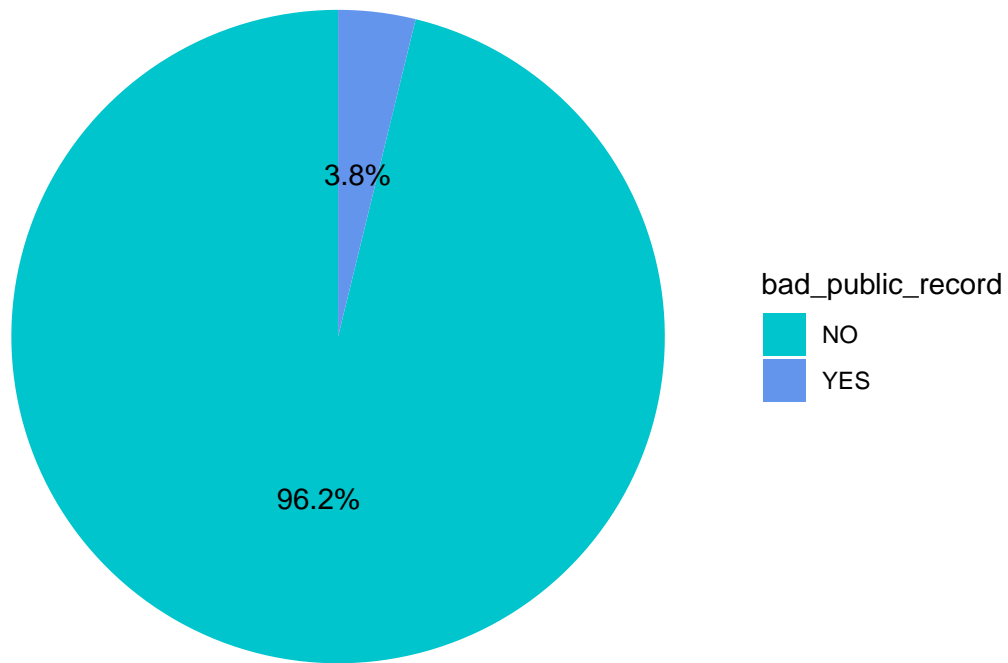
Analisis Univariante

##	Record Publico	Frecuencia Absoluta	Frecunecia relativa
## 1	NO	10713	0.96184234
## 2	YES	425	0.03815766

Interpretacion: Con respecto a la tabla de los Record publicos registrados, el 96,2% de los prestatarios no cuentan con un mal registro, pero el 3,8% de los prestatarios tiene un registro negativo acorde al nivel del monto del prestamo del que se tiene registro.



Porcentajes de los Record Publicos registrados con nivel alto de credito



Interpretacion: Visualizando los graficos, se puede afirmar que, para la muestra actual de Registro publico, existe una mayor cantidad de prestatarios con creditos altos que no cuentan con un mal registro, en relacion con el total de 11138 prestatarios

Análisis Descriptivo

A partir de los graficos observados anteriormente es posible deducir que los prestatarios que no cuentan con un registro publico negativo tienen una mayor proporcion en comparacion con los que cuentan con alguno.

Un registro público negativo se refiere a un registro oficial o legal que documenta información desfavorable o adversa sobre una persona, entidad o propiedad. En finanzas un registro publico negativo es la declaracion de Bancarrota.

Por lo tanto se analizara tambien variables que influyen en esta proporcion, ya que la base de datos al estar constituida en su totalidad por variables cualitativas no es posible realizar un estudio de medidas de dispersion.

##	Record Publico	Bancarrota	Frecuencia Absoluta	Frecunecia relativa
## 1	NO	NO	10713	0.96184234
## 2	YES	NO	99	0.00888849
## 3	NO	YES	0	0.00000000
## 4	YES	YES	326	0.02926917

Revisando la tabla se observa que los prestatarios que no tienen un registro publico negativo ni se han declarado en bancarrota es el 96.2% de la muestra, sin embargo hay un 0.8% de personas que si tienen un record publico negativo pero que no se han declarado en bancarrota por lo que se consideraria un valor atipico

Otra de las variables que podrian influir significativamente en la proporcion de los datos es el nivel de ingreso y la puntuacion de credito ya que son factores importantes al momento de evaluar a una persona si esta calificada o esta sujeto a una solicitud de prestamo.

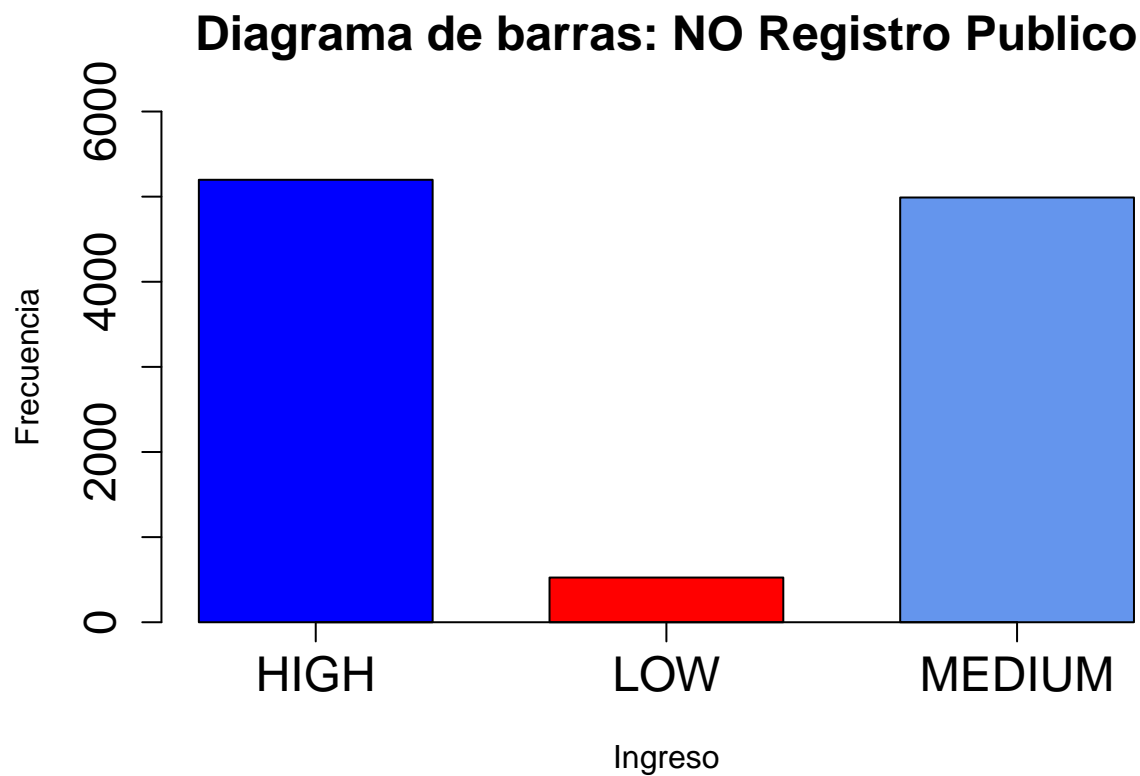
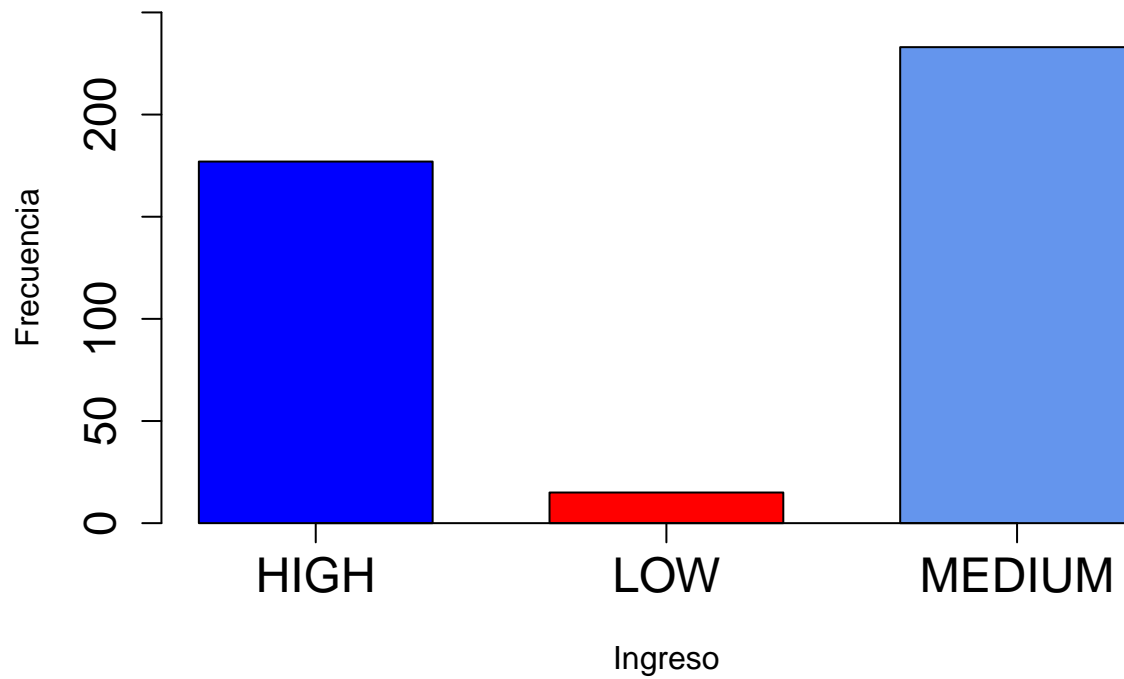
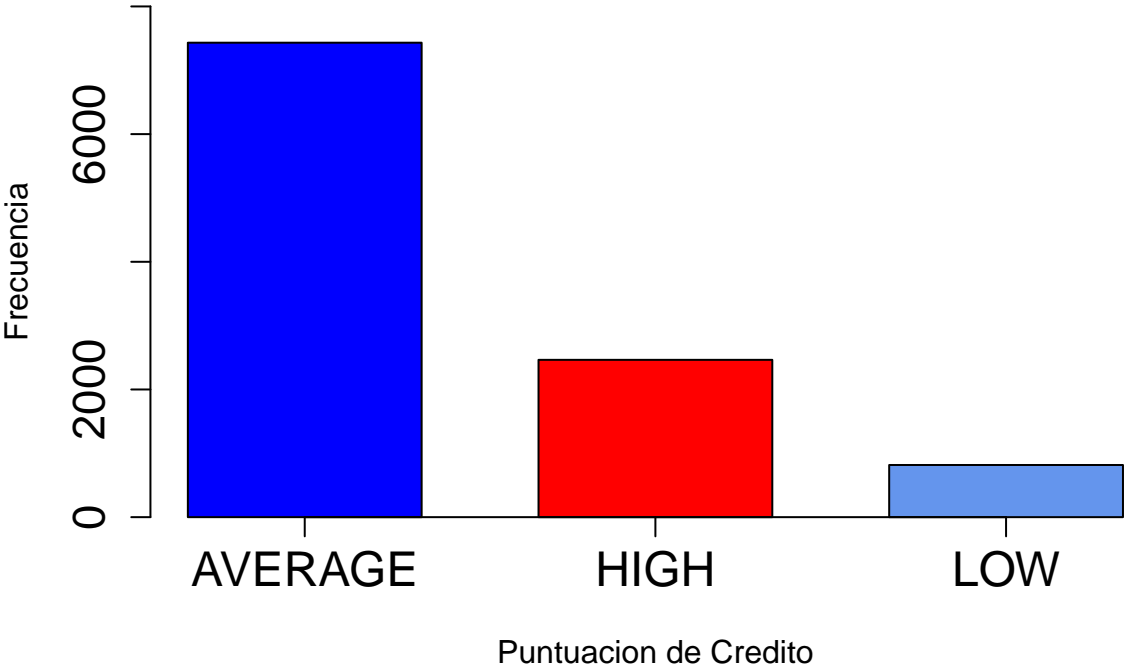


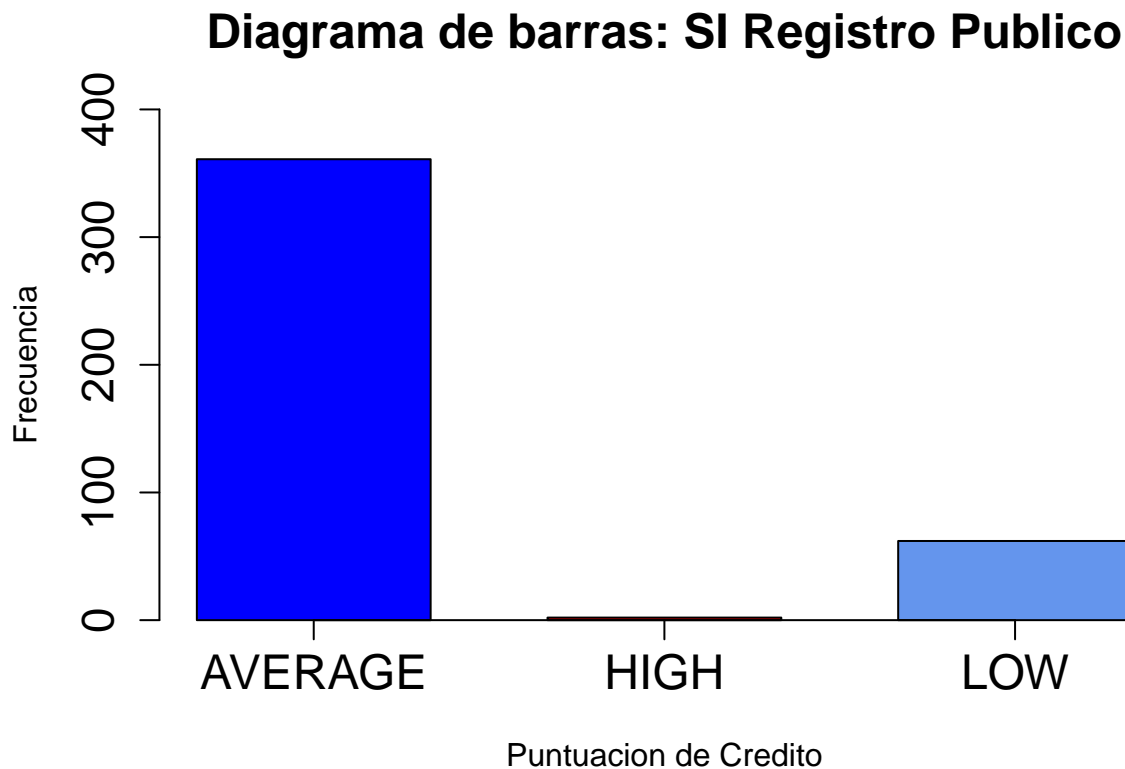
Diagrama de barras: SI Registro Publico



Interpretacion: Como se logra distinguir, los niveles mas destacados entre ambas muestras son HIGH y MEDIUM dando la informacion de que los prestatarios tienen el nivel de ingreso suficiente para ser sujetos de credito.

Diagrama de barras: NO Record Publico





Interpretacion: Podemos notar que el nivel “Average” es el mas alto en ambas graficas, de los prestatarios que NO tienen un registro publico el 69,4% de los 10713 tienen un puntaje de credito promedio, mientras los que SI el 85% de los 425 cuentan con el mismo nivel.

Con estos datos se puede destacar que aunque los prestatarios tenga un registro publico negativo no quiere decir que los ingresos o la puntuacion de credito sea bajo, las graficas muestran que se mantienen entre los niveles altos y medios los cuales influyen en la proporcion final que se desea obtener entre el nivel credito y los registro negativos.

Análisis Inferencial: Contrastar la proporción del monto de préstamo de nivel alto entre los prestatarios que registran un record público negativo y aquellos que no.

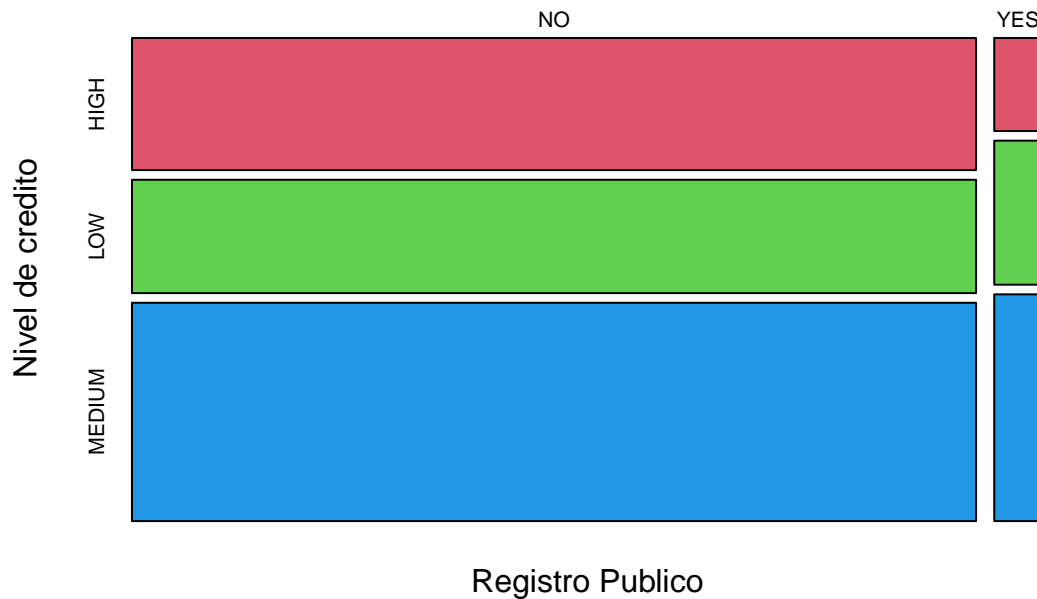
Verificacion de independencia

```
##
## Chi-squared test for given probabilities
##
## data: fa_tabla_PR
## X-squared = 9502.9, df = 1, p-value < 2.2e-16
```

Interpretacion: Se puede observar que mediante la prueba chi cuadrado el valor p es mucho menor a 0.05 se puede decir que existe evidencia suficiente para que las variables sean dependientes una de la otra.

```
td = table(datos$bad_public_record,datos$loan_amount)
mosaicplot(td, main = "Grafico Mosaico: Registro Publico", color = 2:6, xlab = "Registro Publico", ylab = "Loan Amount")
```

Grafico Mosaico: Registro Publico



Interpretacion: En la ilustracion de mosaico, los rectangulos representan la proporcion de las variables, se observa que los prestatarios que no tienen un registro publico negativo tienen un mayor porcentaje y proporcion contrastando con los que si tienen un registro publico negativo.

Hipótesis Nula (H0): La proporción de Registro publicos negativos con nivel de credito "ALTO" es igual a la proporción de Registro publicos negativos con nivel de credito "BAJO"

Hipótesis Alternativa (HA): La proporción de Registro publicos negativos con nivel de credito "ALTO" es diferente a la proporción de Registro publicos negativos con nivel de credito "BAJO"

```
total_PR <- nrow(data_YPR) + nrow(data_NPR)
```

```
prop_NPR <- nrow(data_NPR) / total_PR
```

```
prop_YPR <- nrow(data_YPR) / total_PR
```

```
prop.test(c(nrow(data_NPR), nrow(data_YPR)),
          c(total_PR, total_PR),
          alternative = "two.sided")
```

```
##
```

```
## 2-sample test for equality of proportions with continuity correction
```

```
##
```

```
## data: c(nrow(data_NPR), nrow(data_YPR)) out of c(total_PR, total_PR)
```

```
## X-squared = 19002, df = 1, p-value < 2.2e-16
```

```
## alternative hypothesis: two.sided
```

```
## 95 percent confidence interval:
```

```
## 0.9185633 0.9288060
```

```
## sample estimates:
```

```
## prop 1 prop 2
```

0.96184234 0.03815766

Conclusiones

- **Conclusión 1:** Se concluye que existe dependencia entre las variables y factores que pueden afectar la proporción, es decir, los ingresos y la puntuación de crédito mientras sean altas un prestatario con registro publico negativo puede ser sujeto a un nivel de préstamo alto
- **Conclusión 2:** Se contrastó la proporción entre los registro de los prestatarios y el nivel de préstamo, debido a que el porcentaje destacado mayor recae sobre aquellos que no cuentan con un registro, mientras que el porcentaje más bajo hace referencia a los que si.
- **Conclusión 3:** En base a la información mostrada de los prestatarios, se infiere que tener un registro publico negativo no es una condición suficiente para no solicitar un nivel de préstamo alto

Referencias

Bibliografía

FERNÁNDEZ, J. A. (10 de Enero de 2022). elpais. Obtenido de <https://elpais.com/economia/estar-donde-estes/2022-01-10/lo-que-el-banco-analiza-antes-de-conceder-una-hipoteca-a-un-cliente.html>

Villar, A. (7 de Octubre de 2014). clearpoint. Obtenido de <https://www.clearpoint.org/es/blog/factores-que-influyen-en-su-credito/>

Anexos

<https://github.com/SebastianChiriboga1/Proyecto-Estadistica.git>

Enlace de Repositorio github donde se encuentra el codigo, informe y presentacion del proyecto