Crowfunding en el Mercado de Capitales Colombiano - Proyecto de Estadística para la Analítica de Datos

Sergio Castañeda, Guillermo Castillo, Alexander Guecha

24/04/2022

Justificación: el presente analisis estadístico propone estimar el monto a invertir por un inversionista, teniendo en cuenta las variables que componen la base con la información recopilada de su interacción con el producto.

Con la anterior información se puede ser mas especifico con el portafolio que se le ofrezca a un inversionistas actual y potencial, de cara a generar una diversificación del uso de las alternativas de inversión en personas que han tomado cercanía con la oferta de valor de a2censo.

BDa2censo <- read\_excel("Base de datos inversionistas depurada v3.xlsx",   
 col\_types = c("text", "text", "text",   
 "text", "text", "text", "text", "text",   
 "text", "text", "text", "text", "numeric",   
 "numeric", "text", "text", "date",   
 "numeric", "text", "text", "numeric",   
 "text", "text", "date", "text", "text",   
 "text", "numeric", "text"))

# ANÁLISIS ESTADÍSTICO BASE DE DATOS INVERSIONISTAS DE A2CENSO

## Resumen univariado de variables cualitativas

### Variable 1: SectorCampaña

SectorCampaña\_= BDa2censo$SectorCampaña  
SCtabla=data.frame(table(SectorCampaña\_))  
porcentaje=prop.table(SCtabla[,2])  
SCtabla2= cbind(SCtabla, porcentaje)  
cum\_frequencia=cumsum(SCtabla2[,2])  
SCtabla3= cbind(SCtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(SCtabla3[,3])  
SCtabla4= cbind(SCtabla3, cum\_porcentaje)  
knitr::kable(  
 (SCtabla4)  
)

| SectorCampaña\_ | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| Comercio | 1645 | 0.0813511 | 1645 | 0.0813511 |
| Construcción y Materiales | 1235 | 0.0610751 | 2880 | 0.1424262 |
| Educacion | 116 | 0.0057366 | 2996 | 0.1481628 |
| Energías | 1967 | 0.0972751 | 4963 | 0.2454379 |
| Entretenimiento | 133 | 0.0065773 | 5096 | 0.2520152 |
| Logistica | 915 | 0.0452500 | 6011 | 0.2972652 |
| Movilidad y transporte | 785 | 0.0388210 | 6796 | 0.3360862 |
| Publicidad | 105 | 0.0051926 | 6901 | 0.3412789 |
| RSE | 624 | 0.0308590 | 7525 | 0.3721379 |
| Salud | 2765 | 0.1367390 | 10290 | 0.5088769 |
| Servicios Alimenticios | 4613 | 0.2281292 | 14903 | 0.7370061 |
| Servicios especializados | 287 | 0.0141932 | 15190 | 0.7511992 |
| Tecnología | 3708 | 0.1833737 | 18898 | 0.9345730 |
| Telecomunicaciones | 1130 | 0.0558825 | 20028 | 0.9904555 |
| Turismo | 193 | 0.0095445 | 20221 | 1.0000000 |

barplot(table(BDa2censo$SectorCampaña), col = "blue", main="Diagrama de barras variable Sector Campaña",ylim=c(0,5000),las=2, cex.names = 0.7, lwd = 1.5, cex.axis=0.9 , ylab = "Conteos")

Gráfico, Histograma

Descripción generada automáticamente

plot\_ly(SCtabla4, labels = SCtabla4[,1], values = SCtabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Sector Campaña")

Gráfico, Gráfico circular

Descripción generada automáticamente

Podemos evidenciar que las campañas de inversión que mayor tracción de inversionistas tuvieron fueron de los sectores de servicios alimenticios, tecnología y salud, por otro lado las campañas de sectores económicos de educación, entretenimiento y publicidad fueron las que menos tracción de numero de inversionistas tuvieron.

### Variable 2: Calidad Tributaria

CalidadTributaria= BDa2censo$CalidadTributaria  
CTtabla=data.frame(table(CalidadTributaria))  
porcentaje=prop.table(CTtabla[,2])  
CTtabla2= cbind(CTtabla, porcentaje)  
cum\_frequencia=cumsum(CTtabla2[,2])  
CTtabla3= cbind(CTtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(CTtabla3[,3])  
CTtabla4= cbind(CTtabla3, cum\_porcentaje)  
knitr::kable(  
 (CTtabla4)  
)

| CalidadTributaria | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| AUTO RETENEDOR | 220 | 0.0108798 | 220 | 0.0108798 |
| DECLARANTE | 12068 | 0.5968053 | 12288 | 0.6076851 |
| GRAN CONTRIBUYENTE | 204 | 0.0100885 | 12492 | 0.6177736 |
| NO CONTRIBUYENTE | 6104 | 0.3018644 | 18596 | 0.9196380 |
| NO DECLARANTE | 1625 | 0.0803620 | 20221 | 1.0000000 |

barplot(table(BDa2censo$CalidadTributaria), col = "blue", main="Diagrama de barras",ylim=c(0,13000),las=1, cex.names = 0.5, lwd = 1.5, cex.axis=0.9, xlab="Calidad Tributaria", ylab = "Conteos")

Gráfico, Gráfico de barras

Descripción generada automáticamente

plot\_ly(CTtabla4, labels = CTtabla4[,1], values = CTtabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Calidad Tributaria")

Gráfico, Gráfico circular

Descripción generada automáticamente

Si bien esta variable no es tan relevante para el analisis de la investigación podemos encontrar que la mayor proporsion de inversionistas tienen una calidad tributaria de declarante y no contribuyente y hay una proporsion de un 1% de inversionistas que son grandes contribuyentes.

### Variable 3: Departamento

Departamento= BDa2censo$Departamento  
Dotabla=data.frame(table(Departamento))  
porcentaje=prop.table(Dotabla[,2])  
Dotabla2= cbind(Dotabla, porcentaje)  
cum\_frequencia=cumsum(Dotabla2[,2])  
Dotabla3= cbind(Dotabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Dotabla3[,3])  
Dotabla4= cbind(Dotabla3, cum\_porcentaje)  
knitr::kable(  
 (Dotabla4)  
)

| Departamento | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| Antioquia | 3370 | 0.1666584 | 3370 | 0.1666584 |
| Arauca | 3 | 0.0001484 | 3373 | 0.1668068 |
| Atlantico | 303 | 0.0149844 | 3676 | 0.1817912 |
| Bogotá | 12122 | 0.5994758 | 15798 | 0.7812670 |
| Bolivar | 150 | 0.0074180 | 15948 | 0.7886850 |
| Boyaca | 213 | 0.0105336 | 16161 | 0.7992186 |
| Caldas | 236 | 0.0116710 | 16397 | 0.8108897 |
| Caqueta | 15 | 0.0007418 | 16412 | 0.8116315 |
| Casanare | 20 | 0.0009891 | 16432 | 0.8126205 |
| Cauca | 74 | 0.0036596 | 16506 | 0.8162801 |
| Cesar | 47 | 0.0023243 | 16553 | 0.8186044 |
| Cordoba | 73 | 0.0036101 | 16626 | 0.8222145 |
| Cundinamarca | 916 | 0.0452994 | 17542 | 0.8675140 |
| Huila | 77 | 0.0038079 | 17619 | 0.8713219 |
| La Guajira | 5 | 0.0002473 | 17624 | 0.8715692 |
| Magdalena | 23 | 0.0011374 | 17647 | 0.8727066 |
| Meta | 94 | 0.0046486 | 17741 | 0.8773552 |
| Narino | 78 | 0.0038574 | 17819 | 0.8812126 |
| Norte de Santander | 129 | 0.0063795 | 17948 | 0.8875921 |
| Putumayo | 1 | 0.0000495 | 17949 | 0.8876416 |
| Quindio | 189 | 0.0093467 | 18138 | 0.8969883 |
| Risaralda | 250 | 0.0123634 | 18388 | 0.9093517 |
| San Andres y Providencia | 2 | 0.0000989 | 18390 | 0.9094506 |
| Santander | 597 | 0.0295238 | 18987 | 0.9389743 |
| Sucre | 19 | 0.0009396 | 19006 | 0.9399140 |
| Tolima | 116 | 0.0057366 | 19122 | 0.9456506 |
| Valle del Cauca | 1099 | 0.0543494 | 20221 | 1.0000000 |

barplot(table(BDa2censo$Departamento), col = "blue", main="Diagrama de barras",ylim=c(0,13000),las=2, cex.names = 0.5, lwd = 1.5, cex.axis=0.9, xlab="Departamento", ylab = "Conteos")

Gráfico, Histograma

Descripción generada automáticamente

plot\_ly(Dotabla4, labels = Dotabla4[,1], values = Dotabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Departamento")

Gráfico, Gráfico circular

Descripción generada automáticamente

De otro lado, si bien es esperado que la mayor participación de inversionistas se encuentren ubicados en Bogotá, resulta llamativo observar las participaciones que se tienen en otros departamentos como Antioquia , Valle del cauca, cundinamarca, santander y atlantico.

### Variable 4: Sector Económico del inversionista

SectorEconomico= BDa2censo$SectorEconomico  
SEtabla=data.frame(table(SectorEconomico))  
porcentaje=prop.table(SEtabla[,2])  
SEtabla2= cbind(SEtabla, porcentaje)  
cum\_frequencia=cumsum(SEtabla2[,2])  
SEtabla3= cbind(SEtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(SEtabla3[,3])  
SEtabla4= cbind(SEtabla3, cum\_porcentaje)  
knitr::kable(  
 (SEtabla4)  
)

| SectorEconomico | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| ACUEDUCTO | 6 | 0.0002995 | 6 | 0.0002995 |
| ADMINISTRACIÓN PÚBLICA Y DEFENSA | 76 | 0.0037939 | 82 | 0.0040935 |
| AGROINDUSTRIA | 141 | 0.0070387 | 223 | 0.0111322 |
| ALOJAMIENTO Y SERVICIOS DE COMIDA | 93 | 0.0046426 | 316 | 0.0157748 |
| AUTOMOTRIZ | 40 | 0.0019968 | 356 | 0.0177716 |
| COMERCIO AL POR MAYOR | 115 | 0.0057408 | 471 | 0.0235124 |
| COMERCIO AL POR MENOR | 420 | 0.0209665 | 891 | 0.0444788 |
| CONSTRUCCIÓN | 239 | 0.0119309 | 1130 | 0.0564097 |
| EDUCACIÓN | 626 | 0.0312500 | 1756 | 0.0876597 |
| FINANCIERO | 283 | 0.0141274 | 2039 | 0.1017871 |
| INDUSTRÍAS CREATIVAS Y CULTURALES | 124 | 0.0061901 | 2163 | 0.1079772 |
| INFORMACIÓN Y COMUNICACIONES | 1824 | 0.0910543 | 3987 | 0.1990315 |
| INMOBILIARIAS | 147 | 0.0073383 | 4134 | 0.2063698 |
| INVESTIGACIÓN Y CIENCIAS | 3683 | 0.1838558 | 7817 | 0.3902256 |
| MANTENIMIENTO DE EQUIPOS | 55 | 0.0027456 | 7872 | 0.3929712 |
| MANTENIMIENTO Y CUIDADO TEXTIL | 7 | 0.0003494 | 7879 | 0.3933207 |
| MANUFACTURA | 399 | 0.0199181 | 8278 | 0.4132388 |
| MINERÍA | 191 | 0.0095347 | 8469 | 0.4227736 |
| OTROS SERVICIOS | 78 | 0.0038938 | 8547 | 0.4266673 |
| PERSONA NATURAL | 9731 | 0.4857728 | 18278 | 0.9124401 |
| SALUD | 575 | 0.0287041 | 18853 | 0.9411442 |
| SEGUROS DE SERVICIOS SOCIALES DE RIESGOS PROFESIONALES | 1 | 0.0000499 | 18854 | 0.9411941 |
| SERVICIOS ADMINISTRATIVOS | 838 | 0.0418331 | 19692 | 0.9830272 |
| SERVICIOS DOMÉSTICOS | 28 | 0.0013978 | 19720 | 0.9844249 |
| SERVICIOS ENERGÉTICOS | 52 | 0.0025958 | 19772 | 0.9870208 |
| TRANSPORTE Y ALMACENAMIENTO | 260 | 0.0129792 | 20032 | 1.0000000 |

barplot(table(BDa2censo$SectorEconomico), col = "blue", main="Diagrama de barras variable Sector Economico del Inversionista",ylim=c(0,10000),las=2, cex.names = 0.5, lwd = 1.5, cex.axis=0.9, ylab = "Conteos")

Gráfico

Descripción generada automáticamente

plot\_ly(SEtabla4, labels = SEtabla4[,1], values = SEtabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Sector Económico del Inversionista")

Gráfico, Gráfico circular

Descripción generada automáticamente

La mayor proporsion de inversionistas autodenominan su actividad económica como persona natural, siendo que es la opcion mas entendible para usuarios que no conocen estas clasificaciones económicas, no obtante resulta llamativo observar la participacion de personas que se dedican a actividades de investigación y ciencias, información y comunicaciones y servicios administrativos en donde podemos encontrar posiblemente un proporsion de asalariados.

### Variable 5: Tipo Inversionista

TipoInversionista= BDa2censo$TipoInversionista  
TItabla=data.frame(table(TipoInversionista))  
porcentaje=prop.table(TItabla[,2])  
TItabla2= cbind(TItabla, porcentaje)  
cum\_frequencia=cumsum(TItabla2[,2])  
TItabla3= cbind(TItabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(TItabla3[,3])  
TItabla4= cbind(TItabla3, cum\_porcentaje)  
knitr::kable(  
 (TItabla4)  
)

| TipoInversionista | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| CALIFICADO | 57 | 0.0028189 | 57 | 0.0028189 |
| NO CALIFICADO | 20164 | 0.9971811 | 20221 | 1.0000000 |

barplot(table(BDa2censo$TipoInversionista), col = "blue", main="Diagrama de barras",ylim=c(0,21000),las=2, cex.names = 0.5, lwd = 1.5, cex.axis=0.9, xlab="Tipo Inversionista", ylab = "Conteos")

Gráfico, Forma

Descripción generada automáticamente

plot\_ly(TItabla4, labels = TItabla4[,1], values = TItabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Tipo Inversionista")

Gráfico, Gráfico circular

Descripción generada automáticamente

Por otro lado, podemos encontrar que la proporsion de inversionistas calificados que han invertido en a2censo es muy baja, representando una oportunidad de abordar más a este nicho de inversionistas con este producto.

### Variable 6: Propósito del inversionista en A2censo

Proposito = BDa2censo$Propósito  
PTtabla=data.frame(table(Proposito))  
porcentaje=prop.table(PTtabla[,2])  
PTtabla2= cbind(PTtabla, porcentaje)  
cum\_frequencia=cumsum(PTtabla2[,2])  
PTtabla3= cbind(PTtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(PTtabla3[,3])  
PTtabla4= cbind(PTtabla3, cum\_porcentaje)  
knitr::kable(  
 (PTtabla4)  
)

| Proposito | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| Apoyar el crecimiento de empresas colombianas | 2521 | 0.1246724 | 2521 | 0.1246724 |
| Aprender de financiación | 704 | 0.0348153 | 3225 | 0.1594877 |
| Diversificar mi portafolio de inversión | 7929 | 0.3921171 | 11154 | 0.5516048 |
| Hacer crecer mi dinero | 8730 | 0.4317294 | 19884 | 0.9833342 |
| Hacer parte de la comunidad a2censo - | 337 | 0.0166658 | 20221 | 1.0000000 |

barplot(table(BDa2censo$Propósito), col = "blue", main="Diagrama de barras variable propósito del inversionista en A2censo",ylim=c(0,10000),las=2, cex.names = 0.6, lwd = 1.5, cex.axis=0.9, ylab = "Conteos")

Gráfico, Histograma

Descripción generada automáticamente

plot\_ly(PTtabla4, labels = PTtabla4[,1], values = PTtabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por propósito del inversionista en A2censo")

Gráfico, Gráfico circular

Descripción generada automáticamente

De acuerdo a los anteriores analisis podemos evidenciar como la mayoria de inversionistas actuales siguen propósitos de inversión asociados al crecimiento y rentabilidad del dinero y a la diversificación del portafolio de inversion. No obstante podemos encontrar que hay un 12.5% de la poblacion que tienen fines filantrópicos de apoyar el emprendimiento Colombiano.

### Variable 7: Grupo de Edad

GrupoEdad = BDa2censo$GrupoEdad  
GEtabla=data.frame(table(GrupoEdad))  
porcentaje=prop.table(GEtabla[,2])  
GEtabla2= cbind(GEtabla, porcentaje)  
cum\_frequencia=cumsum(GEtabla2[,2])  
GEtabla3= cbind(GEtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(GEtabla3[,3])  
GEtabla4= cbind(GEtabla3, cum\_porcentaje)  
knitr::kable(  
 (GEtabla4)  
)

| GrupoEdad | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| Entre 18 y 24 | 375 | 0.0185451 | 375 | 0.0185451 |
| Entre 25 y 35 | 7783 | 0.3848969 | 8158 | 0.4034420 |
| Entre 35 y 45 | 7650 | 0.3783196 | 15808 | 0.7817615 |
| Entre 45 y 55 | 2592 | 0.1281836 | 18400 | 0.9099451 |
| Entre 55 y 65 | 1087 | 0.0537560 | 19487 | 0.9637011 |
| Mayores a 65 | 734 | 0.0362989 | 20221 | 1.0000000 |

barplot(table(BDa2censo$GrupoEdad), col = "blue", main="Diagrama de barras variable Grupo de Edad",ylim=c(0,10000),las=2, cex.names = 0.6, lwd = 1.5, cex.axis=0.9, ylab = "Conteos")

Gráfico, Histograma

Descripción generada automáticamente

plot\_ly(GEtabla4, labels = GEtabla4[,1], values = GEtabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Grupo de Edad")

Gráfico, Gráfico circular

Descripción generada automáticamente

De acuerdo al anterior analisis, resulta interesante observar la importante participación de inversionistas jovenes entre 25 y 35 años y tambien de inversionistas entre 35 y 45 años. No obstante la participacion de inversionistas entre 45 y 55 años (12,8%) no es una participación despreciable.

### Variable 8: Grupo de Ingresos

GrupoIngresos = BDa2censo$GrupoIngresos  
GItabla=data.frame(table(GrupoIngresos))  
porcentaje=prop.table(GItabla[,2])  
GItabla2= cbind(GItabla, porcentaje)  
cum\_frequencia=cumsum(GItabla2[,2])  
GItabla3= cbind(GItabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(GItabla3[,3])  
GItabla4= cbind(GItabla3, cum\_porcentaje)  
knitr::kable(  
 (GItabla4)  
)

| GrupoIngresos | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| Entre 1M y 3M | 3960 | 0.1958360 | 3960 | 0.1958360 |
| Entre 3M y 5M | 5202 | 0.2572573 | 9162 | 0.4530933 |
| Entre 5M y 7M | 2988 | 0.1477672 | 12150 | 0.6008605 |
| Entre 7M y 9M | 1781 | 0.0880768 | 13931 | 0.6889372 |
| Menores a 1M | 739 | 0.0365462 | 14670 | 0.7254834 |
| Superiores a 10M | 5551 | 0.2745166 | 20221 | 1.0000000 |

barplot(table(BDa2censo$GrupoIngresos), col = "blue", main="Diagrama de barras variable Grupo de Ingresos",ylim=c(0,6000),las=2, cex.names = 0.6, lwd = 1.5, cex.axis=0.9, ylab = "Conteos")

Gráfico, Histograma

Descripción generada automáticamente

plot\_ly(GItabla4, labels = GItabla4[,1], values = GItabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Grupo de Ingresos")

Gráfico, Gráfico circular

Descripción generada automáticamente

Dentro de los inversionistas de a2censo encontramos inversionistas de dos grupos de rangos salariales predominantes, dentro de los cuales se encuentran inversionistas de ingresos medios (3 y 5 millones), Inversionistas de ingresos altos (Superiores a 10 Millones) y más leve participación de inversionistas de ingresos medio - bajos(de 1 a 3 Millones). No obstante podemos observar que la mayor proporsión de inversionistas suelen ser personas con ingresos superiores a 3 millones de pesos.

### Variable 9: Mes

Mes = BDa2censo$Mes  
Mtabla=data.frame(table(Mes))  
porcentaje=prop.table(Mtabla[,2])  
Mtabla2= cbind(Mtabla, porcentaje)  
cum\_frequencia=cumsum(Mtabla2[,2])  
Mtabla3= cbind(Mtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Mtabla3[,3])  
Mtabla4= cbind(Mtabla3, cum\_porcentaje)  
Mtabla4 = Mtabla4[order(Mtabla4$Freq), ]  
knitr::kable(  
 (Mtabla4)  
)

|  | Mes | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 3 | 11 | 703 | 0.0347658 | 4034 | 0.1994956 |
| 10 | 7 | 1092 | 0.0540033 | 16328 | 0.8074774 |
| 6 | 3 | 1165 | 0.0576134 | 9435 | 0.4665941 |
| 1 | 1 | 1497 | 0.0740319 | 1497 | 0.0740319 |
| 11 | 8 | 1665 | 0.0823401 | 17993 | 0.8898175 |
| 5 | 2 | 1727 | 0.0854063 | 8270 | 0.4089808 |
| 9 | 6 | 1740 | 0.0860492 | 15236 | 0.7534741 |
| 2 | 10 | 1834 | 0.0906978 | 3331 | 0.1647297 |
| 8 | 5 | 1923 | 0.0950992 | 13496 | 0.6674250 |
| 7 | 4 | 2138 | 0.1057317 | 11573 | 0.5723258 |
| 12 | 9 | 2228 | 0.1101825 | 20221 | 1.0000000 |
| 4 | 12 | 2509 | 0.1240789 | 6543 | 0.3235745 |

barplot(table(BDa2censo$Mes), col = "blue", main="Diagrama de barras variable Mes de Inversión",ylim=c(0,3000),las=1, cex.names = 0.6, lwd = 1.5, cex.axis=0.9, ylab = "Conteos")

Gráfico, Histograma

Descripción generada automáticamente

plot\_ly(Mtabla4, labels = Mtabla4[,1], values = Mtabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Mes de Inversió")

Gráfico, Gráfico circular

Descripción generada automáticamente

Resulta complejo poder tomar alguna conclusión respecto al mes de inversión teniendo en cuenta que el año 2020 fue el primero año de ingreso del a2censo al mercado y actualmente 2022 es un año en curso. No obstante resulta llamativo ver la baja participación de inversionistas en el mes 11 (noviembre) en donde hay que revisar la disponibilidad de campañas para invertir u otros aspectos que puedan estar relacionados con menor frecuencia de inversionistas.

### Variable 10: Año

Año = BDa2censo$Año  
Atabla=data.frame(table(Año))  
porcentaje=prop.table(Atabla[,2])  
Atabla2= cbind(Atabla, porcentaje)  
cum\_frequencia=cumsum(Atabla2[,2])  
Atabla3= cbind(Atabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Atabla3[,3])  
Atabla4= cbind(Atabla3, cum\_porcentaje)  
Atabla4 = Atabla4[order(Atabla4$Freq), ]  
knitr::kable(  
 (Atabla4)  
)

|  | Año | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- | --- |
| 3 | 2022 | 1983 | 0.0980664 | 20221 | 1.0000000 |
| 1 | 2020 | 4655 | 0.2302062 | 4655 | 0.2302062 |
| 2 | 2021 | 13583 | 0.6717274 | 18238 | 0.9019336 |

barplot(table(BDa2censo$Año), col = "blue", main="Diagrama de barras variable Año de Inversión",ylim=c(0,15000),las=1, cex.names = 0.6, lwd = 1.5, cex.axis=0.9, ylab = "Conteos")

Gráfico, Gráfico de barras

Descripción generada automáticamente

plot\_ly(Atabla4, labels = Atabla4[,1], values = Atabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores por Año de Inversión")

Gráfico, Gráfico circular

Descripción generada automáticamente

Como mencionabamos antiormente podemos observar una mayor proporsión de inversionistas en el año 2021 lo cual se explica puesto que 2021 fue el año de consolidación del producto en el mercado, 2020 por su lado fue un año de consolidación e ingreso del producto en el mercado en donde aun su posicionamiento no era tan fuerte, y para 2022 tenemos una medición parcializada hasta el mes de marzo.

InvRecurr = BDa2censo$CodEsInversionistaRecurrente  
Rtabla=data.frame(table(InvRecurr))  
porcentaje=prop.table(Rtabla[,2])  
Rtabla2= cbind(Rtabla, porcentaje)  
cum\_frequencia=cumsum(Rtabla2[,2])  
Rtabla3= cbind(Rtabla2, cum\_frequencia)  
cum\_porcentaje=cumsum(Rtabla3[,3])  
Rtabla4= cbind(Rtabla3, cum\_porcentaje)  
knitr::kable(  
 (Rtabla4)  
)

| InvRecurr | Freq | porcentaje | cum\_frequencia | cum\_porcentaje |
| --- | --- | --- | --- | --- |
| 0 | 12707 | 0.6284061 | 12707 | 0.6284061 |
| 1 | 7514 | 0.3715939 | 20221 | 1.0000000 |

barplot(table(BDa2censo$CodEsInversionistaRecurrente), col = "blue", main="Diagrama de barras variable ¿es inversionista recurrente?",ylim=c(0,10000),las=2, cex.names = 0.6, lwd = 1.5, cex.axis=0.9, ylab = "Conteos")

Gráfico

Descripción generada automáticamente

plot\_ly(Rtabla4, labels = Rtabla4[,1], values = Rtabla4[,3], type = "pie") %>%  
 layout(title = "Diagrama de sectores ¿Es inversionista recurrente")

Gráfico, Gráfico circular

Descripción generada automáticamente

Podemos evidenciar que la mayoría (63%) de inversionistas no son recurrentes (identificados con el valor 0)

## Resumen univariado de variables cuantitativas

A continuación se presenta el análisis variable por variable en lo que respecta a información cuantitativa.

Vale la pena resaltar que dichos análisis se hicieron siguiendo la regla de establecimiento de clases de Sturges que incluye la función fdt del paquete fdth.

Este proceso se realizó de esta manera puesto que al realizar los análisis bajo un numero entre 3 y 7 intervalos de clase que fueron probados en cada una de las variables, la distribución de la información no permitía desagregarla en grupos que contribuyeran al análisis y que se representaran de mejor manera, en todo caso la distribución en clases bajo la regla de sturges permitió la mejor visualización de la información cuantitativa.

### Variable 11: Ingresos Mes

basicStats(BDa2censo$IngresosMes)

## X..BDa2censo.IngresosMes  
## nobs 2.022100e+04  
## NAs 0.000000e+00  
## Minimum 1.000000e+06  
## Maximum 1.000000e+08  
## 1. Quartile 3.360000e+06  
## 3. Quartile 1.000000e+07  
## Mean 7.880049e+06  
## Median 5.400000e+06  
## Sum 1.593425e+11  
## SE Mean 5.761461e+04  
## LCL Mean 7.767119e+06  
## UCL Mean 7.992978e+06  
## Variance 6.712247e+13  
## Stdev 8.192831e+06  
## Skewness 4.156570e+00  
## Kurtosis 2.848217e+01

IMtabla\_Sturges <- fdt(BDa2censo$IngresosMes,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (IMtabla\_Sturges)  
)

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [990000,7240625) | 12984 | 0.6421047 | 64.2104743 | 12984 | 64.21047 |
| [7240625,13491250) | 4312 | 0.2132437 | 21.3243658 | 17296 | 85.53484 |
| [13491250,19741875) | 1479 | 0.0731418 | 7.3141783 | 18775 | 92.84902 |
| [19741875,25992500) | 892 | 0.0441126 | 4.4112556 | 19667 | 97.26027 |
| [25992500,32243125) | 218 | 0.0107809 | 1.0780871 | 19885 | 98.33836 |
| [32243125,38493750) | 66 | 0.0032639 | 0.3263934 | 19951 | 98.66475 |
| [38493750,44744375) | 125 | 0.0061817 | 0.6181692 | 20076 | 99.28292 |
| [44744375,50995000) | 40 | 0.0019781 | 0.1978142 | 20116 | 99.48074 |
| [50995000,57245625) | 18 | 0.0008902 | 0.0890164 | 20134 | 99.56975 |
| [57245625,63496250) | 26 | 0.0012858 | 0.1285792 | 20160 | 99.69833 |
| [63496250,69746875) | 1 | 0.0000495 | 0.0049454 | 20161 | 99.70328 |
| [69746875,75997500) | 14 | 0.0006923 | 0.0692350 | 20175 | 99.77251 |
| [75997500,82248125) | 19 | 0.0009396 | 0.0939617 | 20194 | 99.86648 |
| [82248125,88498750) | 0 | 0.0000000 | 0.0000000 | 20194 | 99.86648 |
| [88498750,94749375) | 23 | 0.0011374 | 0.1137431 | 20217 | 99.98022 |
| [94749375,1.01e+08) | 4 | 0.0001978 | 0.0197814 | 20221 | 100.00000 |

|  | x |
| --- | --- |
| start | 990000 |
| end | 101000000 |
| h | 6250625 |
| right | 0 |

plot(IMtabla\_Sturges,type='fh',main="Histograma de frecuencias absolutas Ingresos Mes",  
 xlab=" ",ylab="Conteos",col="red", las=2)

Gráfico, Histograma

Descripción generada automáticamente

plot(IMtabla\_Sturges,type='fp',main="Poligono de frecuencias absolutas Ingresos Mes",  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

Gráfico

Descripción generada automáticamente

De acuerdo al análisis de frecuencia podemos encontrar que un total de 12.984 inversionistas se agrupan en salarios entre 990. 000 y 7.240.625 pesos siendo en rango salarial predominante. No obstante el grupo de inversionistas con ingresos entre 7.240.625 y 13.491.250 de pesos representa el 21% de la muestra de inversionistas.

### Variable 12: Patrimonio

basicStats(BDa2censo$Patrimonio)

## X..BDa2censo.Patrimonio  
## nobs 2.022100e+04  
## NAs 0.000000e+00  
## Minimum 1.000000e+07  
## Maximum 1.000000e+10  
## 1. Quartile 5.000000e+07  
## 3. Quartile 3.500000e+08  
## Mean 3.175835e+08  
## Median 1.500000e+08  
## Sum 6.421856e+12  
## SE Mean 4.341007e+06  
## LCL Mean 3.090748e+08  
## UCL Mean 3.260922e+08  
## Variance 3.810515e+17  
## Stdev 6.172937e+08  
## Skewness 8.003262e+00  
## Kurtosis 9.706154e+01

POtabla\_St <- fdt(BDa2censo$Patrimonio,breaks="Sturges",right=F)  
  
knitr::kable(  
 (POtabla\_St)  
)

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [9900000,6.4053125e+08) | 17791 | 0.8798279 | 87.9827902 | 17791 | 87.98279 |
| [6.4053125e+08,1.2711625e+09) | 1573 | 0.0777904 | 7.7790416 | 19364 | 95.76183 |
| [1.2711625e+09,1.9017938e+09) | 450 | 0.0222541 | 2.2254092 | 19814 | 97.98724 |
| [1.9017938e+09,2.532425e+09) | 241 | 0.0119183 | 1.1918303 | 20055 | 99.17907 |
| [2.532425e+09,3.1630563e+09) | 34 | 0.0016814 | 0.1681420 | 20089 | 99.34721 |
| [3.1630563e+09,3.7936875e+09) | 43 | 0.0021265 | 0.2126502 | 20132 | 99.55986 |
| [3.7936875e+09,4.4243188e+09) | 28 | 0.0013847 | 0.1384699 | 20160 | 99.69833 |
| [4.4243188e+09,5.05495e+09) | 6 | 0.0002967 | 0.0296721 | 20166 | 99.72801 |
| [5.05495e+09,5.6855813e+09) | 0 | 0.0000000 | 0.0000000 | 20166 | 99.72801 |
| [5.6855813e+09,6.3162125e+09) | 2 | 0.0000989 | 0.0098907 | 20168 | 99.73790 |
| [6.3162125e+09,6.9468438e+09) | 18 | 0.0008902 | 0.0890164 | 20186 | 99.82691 |
| [6.9468438e+09,7.577475e+09) | 2 | 0.0000989 | 0.0098907 | 20188 | 99.83680 |
| [7.577475e+09,8.2081063e+09) | 0 | 0.0000000 | 0.0000000 | 20188 | 99.83680 |
| [8.2081063e+09,8.8387375e+09) | 1 | 0.0000495 | 0.0049454 | 20189 | 99.84175 |
| [8.8387375e+09,9.4693688e+09) | 17 | 0.0008407 | 0.0840710 | 20206 | 99.92582 |
| [9.4693688e+09,1.01e+10) | 15 | 0.0007418 | 0.0741803 | 20221 | 100.00000 |

|  | x |
| --- | --- |
| start | 9900000 |
| end | 10100000000 |
| h | 630631250 |
| right | 0 |

plot(POtabla\_St,type='fh',main="Histograma de frecuencias absolutas variable Patrimonio",  
 xlab=" ",ylab="Conteos",col="blue", las=2)

Gráfico, Histograma

Descripción generada automáticamente

plot(POtabla\_St,type='fp',main="Poligono de frecuencias absolutas",  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

Imagen que contiene Gráfico

Descripción generada automáticamente

Por otro lado en lo que respecta al patrimonio de los inversionistas podemos encontrar que el 88% de los inversionistas que representan un total de 17.791 individuos tienen patrimonios entre 9.900.000 y 64.053125

### 14: Monto Inversión

basicStats(BDa2censo$MontoInversion)

## X..BDa2censo.MontoInversion  
## nobs 2.022100e+04  
## NAs 0.000000e+00  
## Minimum 2.000000e+05  
## Maximum 5.520000e+07  
## 1. Quartile 3.000000e+05  
## 3. Quartile 2.000000e+06  
## Mean 1.601058e+06  
## Median 7.000000e+05  
## Sum 3.237499e+10  
## SE Mean 2.037790e+04  
## LCL Mean 1.561116e+06  
## UCL Mean 1.641000e+06  
## Variance 8.396949e+12  
## Stdev 2.897749e+06  
## Skewness 6.559907e+00  
## Kurtosis 7.014340e+01

MItabla\_st <- fdt(BDa2censo$MontoInversion,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (MItabla\_st)  
)

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [198000,3670125) | 18089 | 0.8945651 | 89.4565056 | 18089 | 89.45651 |
| [3670125,7142250) | 1443 | 0.0713615 | 7.1361456 | 19532 | 96.59265 |
| [7142250,10614375) | 428 | 0.0211661 | 2.1166114 | 19960 | 98.70926 |
| [10614375,14086500) | 66 | 0.0032639 | 0.3263934 | 20026 | 99.03566 |
| [14086500,17558625) | 63 | 0.0031156 | 0.3115573 | 20089 | 99.34721 |
| [17558625,21030750) | 75 | 0.0037090 | 0.3709015 | 20164 | 99.71811 |
| [21030750,24502875) | 8 | 0.0003956 | 0.0395628 | 20172 | 99.75768 |
| [24502875,27975000) | 13 | 0.0006429 | 0.0642896 | 20185 | 99.82197 |
| [27975000,31447125) | 16 | 0.0007913 | 0.0791257 | 20201 | 99.90109 |
| [31447125,34919250) | 3 | 0.0001484 | 0.0148361 | 20204 | 99.91593 |
| [34919250,38391375) | 1 | 0.0000495 | 0.0049454 | 20205 | 99.92087 |
| [38391375,41863500) | 5 | 0.0002473 | 0.0247268 | 20210 | 99.94560 |
| [41863500,45335625) | 1 | 0.0000495 | 0.0049454 | 20211 | 99.95055 |
| [45335625,48807750) | 1 | 0.0000495 | 0.0049454 | 20212 | 99.95549 |
| [48807750,52279875) | 8 | 0.0003956 | 0.0395628 | 20220 | 99.99505 |
| [52279875,55752000) | 1 | 0.0000495 | 0.0049454 | 20221 | 100.00000 |

|  | x |
| --- | --- |
| start | 198000 |
| end | 55752000 |
| h | 3472125 |
| right | 0 |

plot(MItabla\_st,type='fh',main="Histograma de frecuencias absolutas variable Monto Inversión",  
 xlab=" ",ylab="Conteos",col="blue", las=2)

Gráfico, Histograma

Descripción generada automáticamente

plot(MItabla\_st,type='fp',main="Poligono de frecuencias absolutas",  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

Imagen que contiene Diagrama

Descripción generada automáticamente Respecto al monto invertido podemos encontrar que un 89% de los inversionistas (n= 18.089) invierten en un rango de 198.000 pesos a 3.670.125 pesos. Encontrándose que este es rango del ticket mayoritario de inversión actual.

### Variable: Edad

basicStats(BDa2censo$Edad)

## X..BDa2censo.Edad  
## nobs 2.022100e+04  
## NAs 0.000000e+00  
## Minimum 1.800000e+01  
## Maximum 9.000000e+01  
## 1. Quartile 3.200000e+01  
## 3. Quartile 4.400000e+01  
## Mean 3.956125e+01  
## Median 3.800000e+01  
## Sum 7.999680e+05  
## SE Mean 7.606200e-02  
## LCL Mean 3.941216e+01  
## UCL Mean 3.971034e+01  
## Variance 1.169876e+02  
## Stdev 1.081608e+01  
## Skewness 1.179185e+00  
## Kurtosis 1.491986e+00

Edtabla\_st <- fdt(BDa2censo$Edad,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (Edtabla\_st)  
)

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [17.82,22.39) | 120 | 0.0059344 | 0.5934425 | 120 | 0.5934425 |
| [22.39,26.96) | 1028 | 0.0508382 | 5.0838237 | 1148 | 5.6772662 |
| [26.96,31.52) | 3640 | 0.1800109 | 18.0010880 | 4788 | 23.6783542 |
| [31.52,36.09) | 4372 | 0.2162109 | 21.6210870 | 9160 | 45.2994412 |
| [36.09,40.66) | 3598 | 0.1779338 | 17.7933831 | 12758 | 63.0928243 |
| [40.66,45.23) | 3050 | 0.1508333 | 15.0833292 | 15808 | 78.1761535 |
| [45.23,49.79) | 1358 | 0.0671579 | 6.7157905 | 17166 | 84.8919440 |
| [49.79,54.36) | 1006 | 0.0497503 | 4.9750260 | 18172 | 89.8669700 |
| [54.36,58.93) | 633 | 0.0313041 | 3.1304090 | 18805 | 92.9973790 |
| [58.93,63.5) | 488 | 0.0241333 | 2.4133327 | 19293 | 95.4107116 |
| [63.5,68.06) | 452 | 0.0223530 | 2.2352999 | 19745 | 97.6460116 |
| [68.06,72.63) | 228 | 0.0112754 | 1.1275407 | 19973 | 98.7735522 |
| [72.63,77.2) | 137 | 0.0067751 | 0.6775135 | 20110 | 99.4510657 |
| [77.2,81.77) | 87 | 0.0043025 | 0.4302458 | 20197 | 99.8813115 |
| [81.77,86.33) | 23 | 0.0011374 | 0.1137431 | 20220 | 99.9950546 |
| [86.33,90.9) | 1 | 0.0000495 | 0.0049454 | 20221 | 100.0000000 |

|  | x |
| --- | --- |
| start | 17.8200 |
| end | 90.9000 |
| h | 4.5675 |
| right | 0.0000 |

plot(Edtabla\_st,type='fh',main="Histograma de frecuencias absolutas variable Edad",  
 xlab=" ",ylab="Conteos",col="blue", las=2)

Gráfico, Histograma

Descripción generada automáticamente

plot(Edtabla\_st,type='fp',main="Poligono de frecuencias absolutas Variable Edad",  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

Gráfico, Gráfico de líneas

Descripción generada automáticamente

De acuerdo al análisis cuantitativo de la variable edad podemos encontrar que la mayor cantidad de inversionistas (22%) se encuentran en un rango de edad entre 31 años y 36 años. No obstante la participación de inversionistas entre los 27 años y 31 años también es relevante con un 18% de participación y también el rango comprendido entre los 36 años y 41 años con un 17% de participación y el rango entre 41 años y 45 años con un 15% de la participación total. Dicho lo anterior los inversionistas entre los 26 años y 45 años suman un total del 73% de participación de los inversionistas siendo la edad media 38 años.

### Variable 15 : Cantidad Inversiones

basicStats(BDa2censo$CantidadInversiones)

## X..BDa2censo.CantidadInversiones  
## nobs 2.022100e+04  
## NAs 0.000000e+00  
## Minimum 1.000000e+00  
## Maximum 8.500000e+01  
## 1. Quartile 4.000000e+00  
## 3. Quartile 1.800000e+01  
## Mean 1.286568e+01  
## Median 9.000000e+00  
## Sum 2.601570e+05  
## SE Mean 9.095800e-02  
## LCL Mean 1.268740e+01  
## UCL Mean 1.304397e+01  
## Variance 1.672954e+02  
## Stdev 1.293427e+01  
## Skewness 1.969355e+00  
## Kurtosis 5.195588e+00

CItabla\_st <- fdt(BDa2censo$CantidadInversiones,breaks="Sturges",right=F)  
  
  
knitr::kable(  
 (CItabla\_st)  
)

| Class limits | f | rf | rf(%) | cf | cf(%) |
| --- | --- | --- | --- | --- | --- |
| [0.99,6.2938) | 8148 | 0.4029474 | 40.2947431 | 8148 | 40.29474 |
| [6.2938,11.598) | 3702 | 0.1830770 | 18.3076999 | 11850 | 58.60244 |
| [11.598,16.901) | 2844 | 0.1406459 | 14.0645863 | 14694 | 72.66703 |
| [16.901,22.205) | 2072 | 0.1024677 | 10.2467732 | 16766 | 82.91380 |
| [22.205,27.509) | 1014 | 0.0501459 | 5.0145888 | 17780 | 87.92839 |
| [27.509,32.812) | 738 | 0.0364967 | 3.6496711 | 18518 | 91.57806 |
| [32.812,38.116) | 614 | 0.0303645 | 3.0364473 | 19132 | 94.61451 |
| [38.116,43.42) | 459 | 0.0226992 | 2.2699174 | 19591 | 96.88443 |
| [43.42,48.724) | 78 | 0.0038574 | 0.3857376 | 19669 | 97.27016 |
| [48.724,54.028) | 291 | 0.0143910 | 1.4390980 | 19960 | 98.70926 |
| [54.028,59.331) | 0 | 0.0000000 | 0.0000000 | 19960 | 98.70926 |
| [59.331,64.635) | 180 | 0.0089016 | 0.8901637 | 20140 | 99.59943 |
| [64.635,69.939) | 0 | 0.0000000 | 0.0000000 | 20140 | 99.59943 |
| [69.939,75.242) | 0 | 0.0000000 | 0.0000000 | 20140 | 99.59943 |
| [75.242,80.546) | 0 | 0.0000000 | 0.0000000 | 20140 | 99.59943 |
| [80.546,85.85) | 81 | 0.0040057 | 0.4005737 | 20221 | 100.00000 |

|  | x |
| --- | --- |
| start | 0.99000 |
| end | 85.85000 |
| h | 5.30375 |
| right | 0.00000 |

plot(CItabla\_st,type='fh',main="Histograma de frecuencias absolutas variable Cantidad de inversiones",  
 xlab=" ",ylab="Conteos",col="blue", las=2)

Gráfico, Histograma

Descripción generada automáticamente

plot(CItabla\_st,type='fp',main="Poligono de frecuencias absolutas Variable Cantidad de inversiones",  
 xlab=" ",ylab="Conteos",col="black",lwd=2,las=2)

Gráfico

Descripción generada automáticamente

Respecto a la variable Cantidad de inversiones podemos encontrar que el promedio de inversiones es 9 inversiones. No obstante un 40% (8.148) de los inversionistas han realizado entre 1 y 6 inversiones. un 18% (3.702) han realizado entre 6 y 12 inversiones, un 14% (2.844) han realizado entre 12 y 17 inversiones y un 10% (2.072) han realizado entre 17 y 22 inversiones en la plataforma de a2censo.

## Análisis Bivariado Cualitativo

**Visualización de las variables cualitativas desde la perspectiva de grupo de ingresos**

A continuación el análisis bivariado cualitativo se realizará comparando las variables contra la variable de grupo ingresos

ggplot(BDa2censo,aes(x=BDa2censo$SectorCampaña,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic()

Gráfico, Gráfico de barras

Descripción generada automáticamente

A partir del anterior grafico podemos observar la concentración de inversionistas con ingresos superiores a 10M de pesos en sectores económicos de tecnología y servicios alimenticios, en los cuales la participación de personas con ingresos entre 3M y 5M también es relevante.

Resulta interesante observar que en general las empresas financiadas en a2censo que pertenecen al sector de tecnología Servicios alimenticios , salud y energías, demuestran mayor atracción de inversionistas de alto capital y medio-bajo capital caracterizándose por tener inversionistas con ingresos superiores a 10 millones, y entre 3 y 5 MM

ggplot(BDa2censo,aes(x=BDa2censo$CalidadTributaria,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic()

Gráfico, Gráfico en cascada

Descripción generada automáticamente

No se observa ninguna conclusión relevante respecto a la calidad tributaria

ggplot(BDa2censo,aes(x=BDa2censo$Departamento,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic()

Escala de tiempo

Descripción generada automáticamente con confianza baja

Lejos de encontrar relevante sacar alguna conclusión de acuerdo al rango de ingresos, de manera general podemos observar que departamentos como valle del cauca, Santander, Antioquia son regiones potenciales para atracción de inversionistas en donde a2censo no ha tenido una alta participación a nivel de marketing o acercamiento con inversionistas

ggplot(BDa2censo,aes(x=BDa2censo$CodEsInversionistaRecurrente,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic()

Gráfico, Gráfico de barras

Descripción generada automáticamente

De acuerdo a la anterior grafica podemos observar como los inversionistas recurrentes (Cod 1) son en mayor medida personas con ingresos superiores a 10 M, aun cuando hay una proporción importante de inversionistas recurrentes con ingresos entre 3M y 5M que no son necesariamente de un alto rango de ingresos y aun así están viendo una oportunidad recurrente de invertir en el mercado

En el caso de los inversionistas no recurrentes podemos encontrar como una importante proporción de inversionistas son personas con ingresos entre 1 M a 5 M

ggplot(BDa2censo,aes(x=BDa2censo$GrupoEdad,)) + geom\_bar(aes(fill= BDa2censo$GrupoIngresos)) + coord\_flip() + theme\_classic()

Gráfico, Gráfico de barras

Descripción generada automáticamente

Realizando un análisis bivariado entre el grupo de edad y el rango de ingresos resulta interesante observar la participación que tienen personas entre rangos de edad de 25 a 35 con ingresos entre 1 a 5 millones de pesos los cuales representan un volumen importante, no obstante volúmenes similares se observan el personas con edades entre 35 y 45 años suelen ser en mayor medida personas con ingresos superiores a 10 millones de pesos

**Visualización de las variables cualitativas desde la perspectiva de grupo de edad**

A continuación el análisis bivariado cualitativo se realizará comparando las variables contra la variable de grupo de edad.

ggplot(BDa2censo,aes(x=BDa2censo$SectorCampaña,)) + geom\_bar(aes(fill= BDa2censo$GrupoEdad)) + coord\_flip() + theme\_classic()

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

## Warning: Use of `BDa2censo$SectorCampaña` is discouraged. Use `SectorCampaña`  
## instead.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Podemos observar que de manera general independiente al sector económico de la empresa financiada, los grupos poblacionales que más invierten en a2censo se encuentran entre 25 y 45 años , no obstante se observa una interesante participación de personas mayores de 65 años en las campañas de tecnología , servicios alimenticios y salud

ggplot(BDa2censo,aes(x=BDa2censo$CalidadTributaria,)) + geom\_bar(aes(fill= BDa2censo$GrupoEdad)) + coord\_flip() + theme\_classic()

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

## Warning: Use of `BDa2censo$CalidadTributaria` is discouraged. Use  
## `CalidadTributaria` instead.

Gráfico, Gráfico en cascada

Descripción generada automáticamente

Como se ha observado hasta el momento la variable Calidad tributaria no tiene mayor análisis

ggplot(BDa2censo,aes(x=BDa2censo$Departamento,)) + geom\_bar(aes(fill= BDa2censo$GrupoEdad)) + coord\_flip() + theme\_classic()

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

## Warning: Use of `BDa2censo$Departamento` is discouraged. Use `Departamento`  
## instead.

Escala de tiempo

Descripción generada automáticamente

No es el grupo poblacional mas grande dentro de la comunidad de inversionistas, no obstante puede ser un grupo de atención las personas entre 45 y 55 años los cuales siguen presentes en Antioquia, valle del cauca

ggplot(BDa2censo,aes(x=BDa2censo$CodEsInversionistaRecurrente,)) + geom\_bar(aes(fill= BDa2censo$GrupoEdad)) + coord\_flip() + theme\_classic()

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Del anterior grafico podemos concluir que para el caso de los inversionistas recurrentes hay una leve proporción mayor de inversionistas en rangos de edades entre 35 y 45 años mientras que esta proporción mayor de inversionistas no recurrentes de centra en personas entre 25 y 35 años, aun cuando ambos grupos de edades son predominantes estos dos grupos de edades

**Visualización de las variables cualitativas desde la perspectiva de inversionista recurrente**

A continuación el analisis bivariado cualitativo se realizará comparando las variables contra la variable de categorización si es un inversionista recurrente o no lo es.

ggplot(BDa2censo,aes(x=BDa2censo$SectorCampaña,)) + geom\_bar(aes(fill= BDa2censo$CodEsInversionistaRecurrente)) + coord\_flip() + theme\_classic()

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

## Warning: Use of `BDa2censo$SectorCampaña` is discouraged. Use `SectorCampaña`  
## instead.

Gráfico, Gráfico en cascada

Descripción generada automáticamente

De manera general se puede observar que la cantidad de inversionistas NO recurrentes es mayor para todos los sectores economicos de las empresas financiadas, es una distribución muy acorde al numero de inversionistas sin considerar si es recurrente o no

ggplot(BDa2censo,aes(x=BDa2censo$Departamento,)) + geom\_bar(aes(fill= BDa2censo$CodEsInversionistaRecurrente)) + coord\_flip() + theme\_classic()

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

## Warning: Use of `BDa2censo$Departamento` is discouraged. Use `Departamento`  
## instead.

Gráfico, Gráfico en cascada

Descripción generada automáticamente

ggplot(BDa2censo,aes(x=BDa2censo$GrupoEdad,)) + geom\_bar(aes(fill= BDa2censo$CodEsInversionistaRecurrente)) + coord\_flip() + theme\_classic()

## Warning: Use of `BDa2censo$CodEsInversionistaRecurrente` is discouraged. Use  
## `CodEsInversionistaRecurrente` instead.

## Warning: Use of `BDa2censo$GrupoEdad` is discouraged. Use `GrupoEdad` instead.

Gráfico, Gráfico en cascada

Descripción generada automáticamente

De la anterior grafica resulta interesante observar que actualmente la cantidad de inversionistas recurrentes son mas en personas entre 35 y 45 años y son aun mas en cantidad respecto a personas entre 25 y 35 años, en donde podemos observar que para este rango de edad entre 25 y 35 años la mayor cantidad de inversionistas son no recurrentes

## 2. Resumen Bivariado entre variables Cualitativas y cuantitativas

A continuación se realizará un analisis Bivariado que contemple una de las variables cualitativas de mayor interés del estudio en cuestión, respecto a una de las variables cuantitativas evaluadas.

# 1. Analisis de las variables desde el monto invertido

ggplot(data = BDa2censo, mapping = aes(x = MontoInversion , y = GrupoIngresos)) + geom\_boxplot()

Gráfico

Descripción generada automáticamente

Del anterior diagrama de cajas y bigotes podemos concluir que aunque desde la correlación lineal simple no se encuentra correlación dada la alta dispersión de los datos que se tienen, a nivel del análisis de la media del monto invertido parece comportarse de manera acorde al rango de ingresos en donde podemos observar que a mayor rango de ingresos la media va incrementando proporcionalmente ( Ojo que no es un dato conclusivo para todos los datos con un buen ajuste, si no un análisis desde medida de tendencia central)

ggplot(data = BDa2censo, mapping = aes(x = IngresosMes , y = GrupoIngresos)) + geom\_boxplot()

Gráfico

Descripción generada automáticamente

De manera similar al anterior grafico podemos encontrar que a) la mayor dispersión de datos respecto a ingresos mes se encuentra en las personas con ingresos superiores a 10 millones de pesos, no obstante desde el análisis de medidas de tendencia central y dispersión podemos observar menores desviaciones estándar en lo que respecto a ingresos menores a 10 millones de pesos

ggplot(data = BDa2censo, mapping = aes(x = Patrimonio , y = GrupoIngresos)) + geom\_boxplot()

Gráfico

Descripción generada automáticamente

De manera similar, la medida de tendencia central es acorde a la conclusión de que un mayor promedio de patrimonio reportado es acorde a un mayor rango de ingresos, no obstante la dispersión no permitiría generalizar esta observación para todos los sujetos encontrándose sujetos con alto patrimonio en ingresos entre 1 y 3 millones

ggplot(data = BDa2censo, mapping = aes(x = Edad , y = GrupoIngresos)) + geom\_boxplot()

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

De acuerdo con el comportamiento que muchas veces se ha observado en las estadísticas Nacionales podemos observar que a mayor edad el promedio de ingresos es mayor encontrándose que el promedio de edad de nuestros inversionistas se encuentra en mayor medida entre los 30 y 40 ños

ggplot(data = BDa2censo, mapping = aes(x = CantidadInversiones , y = GrupoIngresos)) + geom\_boxplot()

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Si bien hay outliers en todos los rangos de ingresos, desde la tendencia central podemos ver que el mayor foco de inversionistas recurrentes se encuentra en personas con ingresos entre 5 a 10 millones de pesos

#2. Analisis desde el tipo de inversionista (Recurrente o no Recurrente)

ggplot(data = BDa2censo, mapping = aes(x = MontoInversion , y = CodEsInversionistaRecurrente)) + geom\_boxplot()

Gráfico

Descripción generada automáticamente

Desde el promedio podemos concluir que los inversionistas NO recurrentes pueden llegar a invertir montos sutilmente mayores, no obstante la dispersión de los datos no permite encontrar dicha relación

ggplot(data = BDa2censo, mapping = aes(x = IngresosMes , y = CodEsInversionistaRecurrente)) + geom\_boxplot()

Gráfico

Descripción generada automáticamente

Los inversionistas recurrentes pueden llegar a tener un promedio de ingresos levemente mayor a los inversionistas no recurrentes

ggplot(data = BDa2censo, mapping = aes(x = Patrimonio , y = CodEsInversionistaRecurrente)) + geom\_boxplot()

Gráfico

Descripción generada automáticamente

Los inversionistas Recurrentes llegan a tener un promedio de patrimonio levemente superior, aun cuando hay personas consideradas inversionistas no recurrentes con un muy alto nivel de patrimonio

ggplot(data = BDa2censo, mapping = aes(x = Edad , y = CodEsInversionistaRecurrente)) + geom\_boxplot()

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Los inversionistas recurrentes tiene un promedio de edad mayor, lo cual puede estar asociado con la siguiente grafica en donde observamos que en este rango de edad dichas inversiones recurrentes pueden estar asociadas con propósitos de inversión como diversificar portafolio, aprender de financiación o sencillamente rentabilidad

ggplot(data = BDa2censo, mapping = aes(x = Edad , y = Propósito)) + geom\_boxplot()

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Aquí podemos observar que personas que quieren hacer parte de la comunidad como propósito de inversión tienen una media superior de edad sin tanta desviación desde los datos

ggplot(data = BDa2censo, mapping = aes(x = CantidadInversiones , y = CodEsInversionistaRecurrente)) + geom\_boxplot()

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

De acuerdo al anterior grafico se encuentra que el promedio de inversiones para un inversionista recurrente ronda las 20 inversiones con algunos individuos que han invertido mas de 60 veces

## 2. Resumen Multivariado entre variables Cualitativas y cuantitativas

A continuación se realizará un análisis Multivariado que contemple dos de las variables cualitativas de mayor interés del estudio en cuestión, respecto a las variables cuantitativas evaluadas.

# Analisis del monto invertido respecto al grupo de edad y el propósito de inversión

ggplot(BDa2censo, aes(Propósito, GrupoEdad, fill= MontoInversion)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1))

Imagen que contiene Diagrama

Descripción generada automáticamente

Del anterior grafico podemos observar cómo personas entre 35 y 45 años con un mayor monto invertido tienen asociado un propósito de diversificación en sus inversiones lo cual puede ser una gran oportunidad en segmentación de la comunicación

# Analisis de la cantidad de inversiones realizadas respecto al grupo de edad y el propósito de inversión

ggplot(BDa2censo, aes(Propósito, GrupoEdad, fill= CantidadInversiones)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="red") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1))

Gráfico, Gráfico en cascada

Descripción generada automáticamente

De Igual forma a como se había mencionado anteriormente resulta interesante observar el grupo poblacional de personas mayores a 45 años con mayor recurrencia de inversión y que tienen un propósito de aprendizaje en temas de financiación, una oportunidad desde el acompañamiento que se le puede dar a este grupo en términos de educación

# Analisis de la cantidad de inversiones realizadas respecto al sector económico de la campaña y el propósito de inversión

ggplot(BDa2censo, aes(Propósito, SectorCampaña , fill= CantidadInversiones)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="red") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1))

Gráfico, Gráfico en cascada

Descripción generada automáticamente

Después de haber visto que el sector económico de educación no tuvo tanta tracción en número de inversionistas totales pero al verlo aquí desde el comportamiento de inversiones recurrentes, resulta interesante poder profundizar con inversionistas el por qué se suele invertir de manera más recurrente por un solo inversionista en campañas de este sector.

Igualmente llama la atención la cantidad de inversiones recurrentes que reciben sectores económicos como Entretenimiento con propósitos de aprendizaje de financiación que como vimos anteriormente están relacionadas con personas de mas de 45 años

ggplot(BDa2censo, aes(GrupoIngresos, GrupoEdad, fill= MontoInversion)) + geom\_tile() +  
 scale\_fill\_gradient(low="white", high="blue") + theme(text = element\_text(size=9), axis.text.x = element\_text(angle=90, hjust=1))

Imagen que contiene Cuadrado

Descripción generada automáticamente

En este grafico podemos contestar a uno de los objetivos de la investigación de cara a la segmentación de los inversionistas de a2censo de cara a llevarlos a otros productos de mayor costo y complejidad, de acuerdo a lo anterior podemos concluir que personas entre 25 y 35 años (Milennials de ingreso medio) con ingresos entre 5 y 7MM tienen mayores montos invertidos, de igual forma que los mas viejos con este mismo rango de ingresos (Gen X de ingresos medios). Por otro lado observar que en mayores rangos de ingresos los mayores montos de inversión se concentran en personas entre 35 y 55 años de edad (milennial y late milennials de altos ingresos.)

# 4. Resumen Bivariado Cuantitativas - Regresiones Lineales simples, multiples y regresión logistica multiple

## Correlaciones Lineales simples entre las variables cuantitativas y su influencia en el monto invertido

cor(x = BDa2censo$Edad, y = BDa2censo$MontoInversion, method = "pearson")

## [1] 0.1168296

cor(x = BDa2censo$IngresosMes, y = BDa2censo$MontoInversion, method = "pearson")

## [1] 0.2279277

cor(x = BDa2censo$Patrimonio, y = BDa2censo$MontoInversion, method = "pearson")

## [1] 0.2272027

cor(x = BDa2censo$CantidadInversiones, y = BDa2censo$MontoInversion, method = "pearson")

## [1] -0.08697955

#cor(x = BDa2censo$CodProposito, y = BDa2censo$MontoInversion, method = "pearson")  
  
#cor(x = BDa2censo$CodProposito, y = BDa2censo$CantidadInversiones, method = "pearson")  
  
#cor(x = BDa2censo$PotencialP2, y = BDa2censo$Edad, method = "pearson")

Se encuentra que desde una correlación lineal simple no hay correlación entre ninguna de las variables a la luz de la variable monto inversión.

## Regresion Lineales simples entre las variables cuantitativas y su influencia en la cantidad de inversiones

cor(x = BDa2censo$Edad, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] 0.158079

cor(x = BDa2censo$IngresosMes, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] 0.1086793

cor(x = BDa2censo$Patrimonio, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] 0.1453263

cor(x = BDa2censo$MontoInversion, y = BDa2censo$CantidadInversiones, method = "pearson")

## [1] -0.08697955

Se encuentra que desde una correlación lineal simple no hay correlación entre ninguna de las variables a la luz de la variable cantidad de inversiones.

#Grafico General de Correlaciones lineales

corrgram(BDa2censo, order=TRUE, lower.panel=panel.shade,  
 upper.panel=panel.pie, text.panel=panel.txt,  
 main="Matriz de Correlaciones General")

Forma

Descripción generada automáticamente

En el anterior grafico podemos observar que aunque dichas correlaciones no son significativas la relación entre las variables Ingresos mes y patrimonio y la relación entre las variables patrimonio y edad son más fuertes respecto a las demás.

### Analisis de correlaciones múltiples

modelotest <- lm(MontoInversion ~ CantidadInversiones + Edad + IngresosMes + Patrimonio, data = BDa2censo )  
  
summary(modelotest)

##   
## Call:  
## lm(formula = MontoInversion ~ CantidadInversiones + Edad + IngresosMes +   
## Patrimonio, data = BDa2censo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11789812 -1084831 -615956 203171 51258352   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.729e+05 7.571e+04 8.888 <2e-16 \*\*\*  
## CantidadInversiones -3.094e+04 1.529e+03 -20.234 <2e-16 \*\*\*  
## Edad 1.561e+04 1.898e+03 8.222 <2e-16 \*\*\*  
## IngresosMes 5.899e-02 2.592e-03 22.754 <2e-16 \*\*\*  
## Patrimonio 7.677e-04 3.546e-05 21.649 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2758000 on 20216 degrees of freedom  
## Multiple R-squared: 0.09399, Adjusted R-squared: 0.09381   
## F-statistic: 524.3 on 4 and 20216 DF, p-value: < 2.2e-16

Predicciones1<-modelotest$fitted.values  
  
library(ggplot2)  
ggplot(data = BDa2censo, aes(x = predict(modelotest),   
 y = abs(rstudent(modelotest))))+  
 geom\_hline(yintercept = 3, color = "grey", linetype = "dashed")+  
 geom\_point(aes(color = ifelse(abs(rstudent(modelotest)) > 2, "red", "black")))+  
 scale\_color\_identity()+  
 labs(title = "DistribuciÃ³n de los residuos estudentizados",   
 x = "PredicciÃ³n modelo",   
 y = "Residuos estudentizados")+  
 theme\_bw() + theme(plot.title = element\_text(hjust = 0.5))

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Se observa un r cuadrado de 0.09 lo cual evidencia un bajo nivel de ajuste en el modelo a la luz de todas la variables cuantitativas relacionadas respecto al Monto de inversión.

modelotest2 <- lm(CantidadInversiones ~ Patrimonio + MontoInversion + Edad + IngresosMes, data = BDa2censo )  
  
summary(modelotest2)

##   
## Call:  
## lm(formula = CantidadInversiones ~ Patrimonio + MontoInversion +   
## Edad + IngresosMes, data = BDa2censo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -35.967 -8.430 -3.648 4.505 67.753   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.192e+00 3.427e-01 18.065 <2e-16 \*\*\*  
## Patrimonio 2.322e-09 1.626e-10 14.283 <2e-16 \*\*\*  
## MontoInversion -6.417e-07 3.171e-08 -20.234 <2e-16 \*\*\*  
## Edad 1.531e-01 8.592e-03 17.814 <2e-16 \*\*\*  
## IngresosMes 1.153e-07 1.193e-08 9.664 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.56 on 20216 degrees of freedom  
## Multiple R-squared: 0.05678, Adjusted R-squared: 0.05659   
## F-statistic: 304.2 on 4 and 20216 DF, p-value: < 2.2e-16

Predicciones1<-modelotest2$fitted.values

Se observa un r cuadrado de 0.05 lo cual evidencia un bajo nivel de ajuste en el modelo a la luz de todas la variables cuantitativas relacionadas respecto a Cantidad Inversiones.

## Analisis de modelo de regresión incluyendo solo las variables que demostraron aportar información al modelo general

step(object = modelotest, direction = "both", trace = 1) ## calculo de criterio de perdida de info

## Start: AIC=599767.6  
## MontoInversion ~ CantidadInversiones + Edad + IngresosMes + Patrimonio  
##   
## Df Sum of Sq RSS AIC  
## <none> 1.5383e+17 599768  
## - Edad 1 5.1442e+14 1.5434e+17 599833  
## - CantidadInversiones 1 3.1153e+15 1.5694e+17 600171  
## - Patrimonio 1 3.5662e+15 1.5739e+17 600229  
## - IngresosMes 1 3.9396e+15 1.5777e+17 600277

##   
## Call:  
## lm(formula = MontoInversion ~ CantidadInversiones + Edad + IngresosMes +   
## Patrimonio, data = BDa2censo)  
##   
## Coefficients:  
## (Intercept) CantidadInversiones Edad   
## 6.729e+05 -3.094e+04 1.561e+04   
## IngresosMes Patrimonio   
## 5.899e-02 7.677e-04

De acuerdo al algoritmo de cálculo de perdida de información del modelo encontramos que la variable Edad es la indica un menor AIC o criterio de perdida de información

names(BDa2censo)

## [1] "Campaña" "SectorCampaña"   
## [3] "Tipo Documento de Identificación" "ClasePersona"   
## [5] "CalidadTributaria" "Pais"   
## [7] "Departamento" "Ciudad"   
## [9] "CIIU" "SectorEconomico"   
## [11] "TipoInversionista" "CodTipoInversionista"   
## [13] "IngresosMes" "Patrimonio"   
## [15] "Propósito" "CodProposito"   
## [17] "FechaNacimiento" "MontoInversion"   
## [19] "NumeroCampaña" "EsUsuarioNuevo?"   
## [21] "Edad" "GrupoEdad"   
## [23] "CodEsInversionistaRecurrente" "FechaCierreCampaña"   
## [25] "Mes" "Año"   
## [27] "CodAño" "CantidadInversiones"   
## [29] "GrupoIngresos"

modeloDef <- lm(MontoInversion ~ CantidadInversiones ,data = BDa2censo )  
  
summary(modeloDef)

##   
## Call:  
## lm(formula = MontoInversion ~ CantidadInversiones, data = BDa2censo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1632280 -1295874 -812793 204885 53367720   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1851766 28635 64.67 <2e-16 \*\*\*  
## CantidadInversiones -19487 1570 -12.41 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2887000 on 20219 degrees of freedom  
## Multiple R-squared: 0.007565, Adjusted R-squared: 0.007516   
## F-statistic: 154.1 on 1 and 20219 DF, p-value: < 2.2e-16

Después de correr varias regresiones, se concluye que no hay una correlación significativa en las variables de manera individual o agrupada y por tanto no hay un modelo que permita predecir el monto de inversión o la cantidad de inversiones realizadas

### Regresión logistica

BDa2censo$CodEsInversionistaRecurrente<-as.factor(BDa2censo$CodEsInversionistaRecurrente)  
BDa2censo$SectorCampaña<-as.factor(BDa2censo$SectorCampaña)  
BDa2censo$CalidadTributaria<-as.factor(BDa2censo$CalidadTributaria)  
BDa2censo$Propósito<-as.factor(BDa2censo$Propósito)  
#BDa2censo$CodProposito<-as.factor(BDa2censo$CodProposito)  
#BDa2censo$PotencialP2<-as.factor(BDa2censo$PotencialP2)

set.seed(88)  
split <- sample.split(BDa2censo$CodEsInversionistaRecurrente, SplitRatio = 0.75)  
training <- subset(BDa2censo,split =="TRUE")  
testing <- subset(BDa2censo,split=="FALSE")

logMultiple=glm(data=training,CodEsInversionistaRecurrente~MontoInversion+Edad+IngresosMes+Patrimonio+SectorCampaña+Propósito, family = binomial)  
summary(logMultiple)

##   
## Call:  
## glm(formula = CodEsInversionistaRecurrente ~ MontoInversion +   
## Edad + IngresosMes + Patrimonio + SectorCampaña + Propósito,   
## family = binomial, data = training)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5701 -0.9538 -0.7936 1.2758 3.0728   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -1.232e+00 1.011e-01 -12.190  
## MontoInversion -1.290e-07 9.389e-09 -13.744  
## Edad 1.982e-02 1.670e-03 11.864  
## IngresosMes 2.378e-08 2.483e-09 9.578  
## Patrimonio 1.780e-10 3.415e-11 5.212  
## SectorCampañaConstrucción y Materiales -2.427e-02 9.061e-02 -0.268  
## SectorCampañaEducacion 8.573e-01 2.324e-01 3.689  
## SectorCampañaEnergías -1.715e-01 8.087e-02 -2.121  
## SectorCampañaEntretenimiento 2.359e-01 2.108e-01 1.119  
## SectorCampañaLogistica 5.975e-02 9.902e-02 0.603  
## SectorCampañaMovilidad y transporte 1.971e-01 1.043e-01 1.890  
## SectorCampañaPublicidad 2.987e-01 2.350e-01 1.271  
## SectorCampañaRSE 1.394e-01 1.117e-01 1.249  
## SectorCampañaSalud -2.579e-01 7.598e-02 -3.394  
## SectorCampañaServicios Alimenticios -2.449e-01 6.959e-02 -3.519  
## SectorCampañaServicios especializados -4.719e-01 1.624e-01 -2.906  
## SectorCampañaTecnología -1.629e-01 7.183e-02 -2.267  
## SectorCampañaTelecomunicaciones 1.406e-01 9.233e-02 1.522  
## SectorCampañaTurismo -4.559e-01 1.908e-01 -2.390  
## PropósitoAprender de financiación -1.268e-01 1.061e-01 -1.195  
## PropósitoDiversificar mi portafolio de inversión 2.067e-01 5.575e-02 3.707  
## PropósitoHacer crecer mi dinero -2.246e-01 5.604e-02 -4.007  
## PropósitoHacer parte de la comunidad a2censo - -4.129e-01 1.464e-01 -2.820  
## Pr(>|z|)   
## (Intercept) < 2e-16 \*\*\*  
## MontoInversion < 2e-16 \*\*\*  
## Edad < 2e-16 \*\*\*  
## IngresosMes < 2e-16 \*\*\*  
## Patrimonio 1.87e-07 \*\*\*  
## SectorCampañaConstrucción y Materiales 0.788850   
## SectorCampañaEducacion 0.000225 \*\*\*  
## SectorCampañaEnergías 0.033928 \*   
## SectorCampañaEntretenimiento 0.263167   
## SectorCampañaLogistica 0.546203   
## SectorCampañaMovilidad y transporte 0.058694 .   
## SectorCampañaPublicidad 0.203615   
## SectorCampañaRSE 0.211715   
## SectorCampañaSalud 0.000689 \*\*\*  
## SectorCampañaServicios Alimenticios 0.000434 \*\*\*  
## SectorCampañaServicios especializados 0.003659 \*\*   
## SectorCampañaTecnología 0.023365 \*   
## SectorCampañaTelecomunicaciones 0.127928   
## SectorCampañaTurismo 0.016867 \*   
## PropósitoAprender de financiación 0.232043   
## PropósitoDiversificar mi portafolio de inversión 0.000210 \*\*\*  
## PropósitoHacer crecer mi dinero 6.15e-05 \*\*\*  
## PropósitoHacer parte de la comunidad a2censo - 0.004807 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 20013 on 15165 degrees of freedom  
## Residual deviance: 19181 on 15143 degrees of freedom  
## AIC: 19227  
##   
## Number of Fisher Scoring iterations: 4

#str(data)

Desde el modelo de regresión logística diseñado se encuentra que el sector económico de la campaña es una variable con un nivel de significancia en la variable de recurrencia del inversionista. De la misma forma en que muestra significancia respecto a los propósitos de inversión de diversificación del portafolio y el motivador de hacer crecer el capital.

De otro lado, las variables monto inversión, edad e ingresos mes también muestran significancia respecto a la variable asociada a si el individuo es inversionista recurrente o no

logMultiple4=glm(data=training,CodEsInversionistaRecurrente~MontoInversion+Edad+IngresosMes+Patrimonio+SectorCampaña+Propósito+CalidadTributaria, family = binomial)  
summary(logMultiple4)

##   
## Call:  
## glm(formula = CodEsInversionistaRecurrente ~ MontoInversion +   
## Edad + IngresosMes + Patrimonio + SectorCampaña + Propósito +   
## CalidadTributaria, family = binomial, data = training)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.6178 -0.9594 -0.7764 1.2561 3.0969   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -1.905e+00 2.234e-01 -8.527  
## MontoInversion -1.320e-07 9.485e-09 -13.918  
## Edad 1.758e-02 1.696e-03 10.364  
## IngresosMes 1.944e-08 2.501e-09 7.773  
## Patrimonio 2.037e-10 3.459e-11 5.890  
## SectorCampañaConstrucción y Materiales -2.187e-02 9.101e-02 -0.240  
## SectorCampañaEducacion 9.022e-01 2.345e-01 3.847  
## SectorCampañaEnergías -1.686e-01 8.130e-02 -2.074  
## SectorCampañaEntretenimiento 2.302e-01 2.129e-01 1.081  
## SectorCampañaLogistica 7.145e-02 9.961e-02 0.717  
## SectorCampañaMovilidad y transporte 1.983e-01 1.049e-01 1.891  
## SectorCampañaPublicidad 2.814e-01 2.356e-01 1.194  
## SectorCampañaRSE 1.493e-01 1.123e-01 1.330  
## SectorCampañaSalud -2.573e-01 7.637e-02 -3.369  
## SectorCampañaServicios Alimenticios -2.422e-01 6.997e-02 -3.462  
## SectorCampañaServicios especializados -4.455e-01 1.634e-01 -2.726  
## SectorCampañaTecnología -1.571e-01 7.222e-02 -2.176  
## SectorCampañaTelecomunicaciones 1.340e-01 9.278e-02 1.444  
## SectorCampañaTurismo -4.184e-01 1.924e-01 -2.175  
## PropósitoAprender de financiación -1.508e-01 1.065e-01 -1.416  
## PropósitoDiversificar mi portafolio de inversión 2.107e-01 5.606e-02 3.758  
## PropósitoHacer crecer mi dinero -1.954e-01 5.641e-02 -3.463  
## PropósitoHacer parte de la comunidad a2censo - -3.594e-01 1.474e-01 -2.439  
## CalidadTributariaDECLARANTE 9.099e-01 2.019e-01 4.506  
## CalidadTributariaGRAN CONTRIBUYENTE -1.094e+00 3.366e-01 -3.249  
## CalidadTributariaNO CONTRIBUYENTE 6.363e-01 2.033e-01 3.130  
## CalidadTributariaNO DECLARANTE 5.059e-01 2.116e-01 2.391  
## Pr(>|z|)   
## (Intercept) < 2e-16 \*\*\*  
## MontoInversion < 2e-16 \*\*\*  
## Edad < 2e-16 \*\*\*  
## IngresosMes 7.67e-15 \*\*\*  
## Patrimonio 3.86e-09 \*\*\*  
## SectorCampañaConstrucción y Materiales 0.810066   
## SectorCampañaEducacion 0.000119 \*\*\*  
## SectorCampañaEnergías 0.038049 \*   
## SectorCampañaEntretenimiento 0.279478   
## SectorCampañaLogistica 0.473180   
## SectorCampañaMovilidad y transporte 0.058649 .   
## SectorCampañaPublicidad 0.232430   
## SectorCampañaRSE 0.183636   
## SectorCampañaSalud 0.000754 \*\*\*  
## SectorCampañaServicios Alimenticios 0.000536 \*\*\*  
## SectorCampañaServicios especializados 0.006407 \*\*   
## SectorCampañaTecnología 0.029583 \*   
## SectorCampañaTelecomunicaciones 0.148601   
## SectorCampañaTurismo 0.029666 \*   
## PropósitoAprender de financiación 0.156918   
## PropósitoDiversificar mi portafolio de inversión 0.000171 \*\*\*  
## PropósitoHacer crecer mi dinero 0.000533 \*\*\*  
## PropósitoHacer parte de la comunidad a2censo - 0.014726 \*   
## CalidadTributariaDECLARANTE 6.59e-06 \*\*\*  
## CalidadTributariaGRAN CONTRIBUYENTE 0.001157 \*\*   
## CalidadTributariaNO CONTRIBUYENTE 0.001748 \*\*   
## CalidadTributariaNO DECLARANTE 0.016783 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 20013 on 15165 degrees of freedom  
## Residual deviance: 19023 on 15139 degrees of freedom  
## AIC: 19077  
##   
## Number of Fisher Scoring iterations: 4

De acuerdo a lo anterior encontramos que el modelo en donde menor perdida de información se da es el modelo 4 en donde el AIC es de 19077 asociado a las variables MontoInversion + Edad + IngresosMes + Patrimonio + SectorCampaña + Propósito + CalidadTributaria)

#PREDICCIONES

res <- predict(logMultiple4, training,type = 'response')

# matriz de confusion

Matriz\_Conf<-table(Predicción=res>0.48, Real=training$CodEsInversionistaRecurrente )  
Matriz\_Conf

## Real  
## Predicción 0 1  
## FALSE 8403 4156  
## TRUE 1127 1480

# Precisión

(Matriz\_Conf[[1,1]]+Matriz\_Conf[[2,2]])/sum(Matriz\_Conf)

## [1] 0.651655

De acuerdo al modelo de regresión logistica determinado podemos encontrar un indice de precisión del 65% con un treshold de 48%

# Curva ROC

test\_prob = predict(logMultiple4, newdata = training, type = "response")  
test\_roc = roc(training$CodEsInversionistaRecurrente ~ test\_prob, plot = TRUE, print.auc = TRUE)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

Gráfico

Descripción generada automáticamente