



FACULTAD DE INGENIERÍA



DIVISIÓN DE INGENIERÍA ELÉCTRICA

Minería de Datos (Clave: 2933)

Proyecto Final || Plan Propuesto

Grupo: 01

Profesor: M.A. Gerardo Gabriel Carrasco Zúñiga

Equipo: 02

Celis Hernández Ronie

Chagoya González Leonardo

López Muñoz José Luis

Nava Alberto Vanessa

Sánchez Wilhelm Jeremías

Índice

Introducción.....	2
Abstract.....	3
Desarrollo del Proyecto.....	4
Información transformada.....	4
Estadísticas.....	7
Aplicación de modelos.....	14
Conclusiones del Proyecto.....	25

Introducción

Este proyecto abarca un planteamiento de una problemática en un banco, donde se busca determinar la mejor zona para colocar una nueva sucursal. Se propondrá una solución estimada para el problema, donde se propondrá una planeación de proyecto y se explicará el desarrollo de dicha solución.

Cabe mencionar que este proyecto tiene como objetivo aplicar los conocimientos adquiridos durante el curso de la materia Minería de Datos, en el cuál se aplicaran conceptos clave como árboles de decisión, clustering, modelo SVM, modelos lineales, entre otros; lo que nos permitirá hacer un análisis de los datos de manera que nos permitan proponer una solución con el fin de poder conocer el comportamiento financiero y social en las alcaldías de la Ciudad de México utilizando herramientas como el software Orange y Excel. Dichas herramientas se pusieron en práctica durante todo el semestre, demostrando los conocimientos adquiridos durante el curso.

Abstract

This project covers an approach to a problem in a bank, which seeks to determine the best area to place a new branch. An estimated solution to the problem will be proposed, where a project planning will be proposed and the development of such solution will be explained.

It is worth mentioning that this project aims to apply the knowledge acquired during the Data Mining course, in which key concepts such as decision trees, clustering, SVM model, linear models, among others, will be applied, which will allow us to analyze the data in a way that will allow us to propose a solution in order to know the financial and social behavior in the municipalities of Mexico City using tools such as Orange and Excel software. These tools were put into practice throughout the semester, demonstrating the knowledge acquired during the course.

Desarrollo del Proyecto

Información transformada

El proceso llevado a cabo consistió en la integración y filtrado de datos para comparar la información de las sucursales bancarias en la Ciudad de México. En primer lugar, se recopilaron los datos de las sucursales de Santander y se realizó una limpieza inicial, contabilizando las sucursales mencionadas en el documento del profesor y asegurándose de que cada sucursal apareciera solo una vez, eliminando así los duplicados. Posteriormente, se aplicó un procedimiento similar a un "join" para identificar las coincidencias entre las sucursales del archivo de minería y las del archivo de Banorte. Dado que el análisis estaba enfocado en delegaciones de la CDMX, se filtraron los datos para mantener sólo aquellas coincidencias en el número de sucursales, descartando las filas que no corresponden.

Elementos de información requeridos

Estos elementos buscan proporcionar una base sólida para determinar la ubicación óptima de una nueva sucursal, permitiendo un análisis exhaustivo del comportamiento financiero de los clientes en las diversas delegaciones de la Ciudad de México. A través del estudio de variables demográficas y financieras, junto con el análisis de la actividad bancaria y la segmentación de clientes por zonas, se espera identificar áreas estratégicas y entender mejor las características y necesidades de la clientela, contribuyendo a una toma de decisiones fundamentada y eficiente.

- **Datos Demográficos y Financieros de Clientes:**
 - Edad, tipo de vivienda (familiar, propia, en renta, o con hipoteca), escolaridad, ingreso inferior y línea de crédito final son los atributos fundamentales.
- **Ubicación y Frecuencia de Clientes en Sucursales:**
 - La ubicación de las sucursales actuales y la frecuencia con la que los clientes las visitan en distintas delegaciones de la Ciudad de México. Esto incluye identificar las delegaciones con mayor o menor actividad bancaria.
- **Estadísticas de Solicitudes y Aprobaciones:**
 - Estadísticas detalladas del número y estatus final de solicitudes de crédito por delegación, que permiten ver las tasas de aprobación y los patrones de rechazo.
- **Comportamiento Financiero de los Clientes por Delegación:**

- Tasa de aprobación/rechazo de solicitudes de crédito en función de variables como el nivel educativo, tipo de vivienda y puntaje de cliente (calificación de riesgo).
- **Segmentación de clientes:**
 - Uso de modelos de clustering para agrupar clientes según su perfil financiero y su potencial de riesgo, evaluando variables como puntaje, ingreso inferido y tipo de vivienda.
- **Relación entre Ingreso y Capacidad de Pago:**
 - Determinar si existe una valoración significativa entre el ingreso inferido de los clientes y su capacidad de pago, así como su comportamiento respecto a las líneas de crédito otorgadas.
- **Análisis Geográfico de Oportunidades para Nuevas Sucursales:**
 - Análisis de zonas estratégicas para una nueva sucursal, basado en el perfil de clientes, frecuencia de visitas y distribución geográfica de la demanda.

Métodos de limpieza y transformación

En esta sección, se describen los métodos de limpieza y transformación de datos utilizados para asegurar la calidad y consistencia de la información, así como su adecuación para los análisis subsiguientes. Estos procesos permiten depurar y estructurar los datos, facilitando el análisis de variables demográficas y financieras, y optimizando su uso en modelos predictivos y de segmentación. Al aplicar técnicas de estandarización, tratamiento de datos faltantes y preparación para el análisis geoespacial, se establece una base sólida para identificar la ubicación óptima de una nueva seguridad.

Identificación de datos relevantes:

- Se seleccionaron las variables críticas para el análisis, como edad, tipo de vivienda, escolaridad, ingreso inferior y ubicación geográfica. Además, se incluyó información sobre frecuencia de visitas y puntuación de cliente, fundamentales para evaluar perfiles financieros.
- **Limpieza de datos:**
 - Se revisaron y eliminaron registros con valores nulos o inconsistentes, particularmente en campos esenciales como ingreso y escolaridad. Los datos erróneos o fuera de

rango se eliminaron o corrigieron, asegurando que cada entrada fuese válida para el análisis.

- **Estandarización de variables:**

- Las categorías de tipo de vivienda y escolaridad fueron estandarizadas para facilitar el análisis, unificando etiquetas como "vivienda familiar" y "propia" en categorías comunes. Se normalizaron los ingresos y el puntaje de cliente para aplicar modelos de análisis.

- **Transformación de datos para análisis geoespacial:**

- Las ubicaciones de las delegaciones y sucursales fueron ajustadas a un formato geoespacial uniforme, permitiendo crear gráficos y mapas que reflejan la frecuencia de visitas y comportamiento por zona.

- **Preparación para modelos analíticos:**

- Los datos se prepararon para su uso en modelos específicos (árbol de decisión, SVM, clustering), transformando variables continuas en categóricas según el requerimiento de cada modelo.

Estadísticas

Una vez que se han realizado los métodos de limpieza y transformación a los datos, se puede realizar un análisis estadístico de aquellas solicitudes que son de la Ciudad de México.

Comenzamos el análisis con una estadística del número de solicitudes realizadas en todas las alcaldías, llegando a un resultado total de 216 solicitudes.

Delegacion	NumSolicitudes
Azcapotzalco	17
Cuajimalpa de Morelos	2
Gustavo A. Madero	12
Iztacalco	4
Iztapalapa	16
Miguel Hidalgo	23
Tlalpan	19
Venustiano Carranza	10
Benito Juarez	27
Cuauhtemoc	24
Alvaro Obregon	18
Coyoacan	44
Total general	216

Figura 1: Tabla de solicitudes de crédito por alcaldía.

De las cuales podemos observar que la delegación que presenta un mayor número de solicitudes es Coyoacán con 44, seguido de Benito Juárez con 27 y Cuauhtémoc con 24. Mientras que en el otro extremo la delegación que menor número de solicitudes tiene es Cuajimalpa de Morelos con apenas 2 solicitudes.

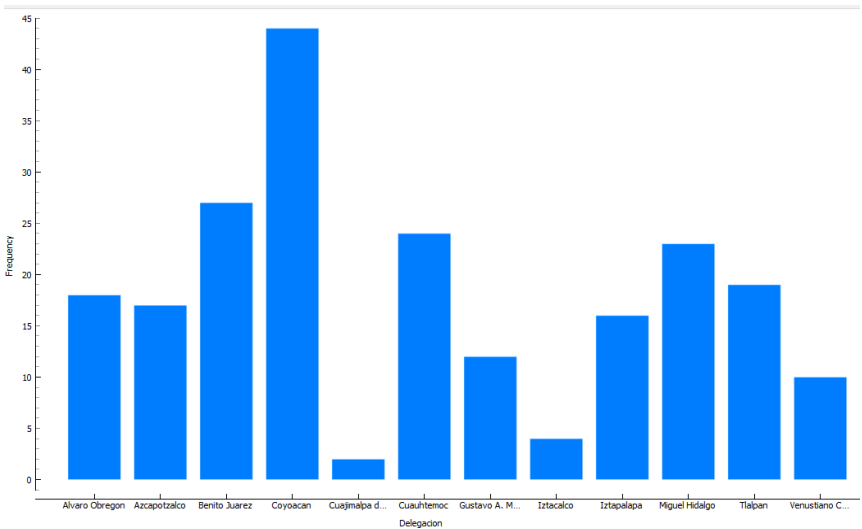


Figura 2: Gráfica de solicitudes de crédito por alcaldía.

Haciendo otro análisis, esta vez con el estatus final de la solicitud de los créditos pudimos encontrar que la alcaldía Coyoacán es la que obtuvo mayor número de solicitudes aceptadas, con un total de 17, seguido de Benito Juárez con 12 y Azcapotzalco con 10.

Delegacion	Aprobado	Pre-Aprobado	Rechazado	Total general	Tasa de aprobacion
Azcapotzalco	8	2	7	17	58.82352941
Cuajimalpa de Morelos	2			2	100
Gustavo A. Madero	1	2	9	12	25
Iztacalco	3		1	4	75
Iztapalapa	4	1	11	16	31.25
Miguel Hidalgo	9		14	23	39.13043478
Tlalpan	3	4	12	19	36.84210526
Venustiano Carranza		2	8	10	20
Benito Juarez	7	5	15	27	44.44444444
Cuauhtemoc	6	2	16	24	33.33333333
Alvaro Obregon	3		15	18	16.66666667
Coyoacan	4	13	27	44	38.63636364
Total general	50	31	135	216	37.5

Figura 3: Tabla de estatus final de solicitudes de crédito por alcaldía.

Ahora, siendo analizado desde la tasa de solicitudes aprobadas, se logra percatar que Cuajimalpa de Morelos aceptó las 2 solicitudes que tuvo, mostrando una tasa de 100% de aprobación, seguido de Azcapotzalco, la cual muestra una tasa de 58.82% y Benito Juárez con 44.44%.

Pero, por otro lado, las delegaciones que presentan una menor tasa de aprobación son Venustiano Carranza con 20%, Gustavo A. Madero con 25% e Iztapalapa con 31.25%

La gráfica se muestra de la siguiente manera:

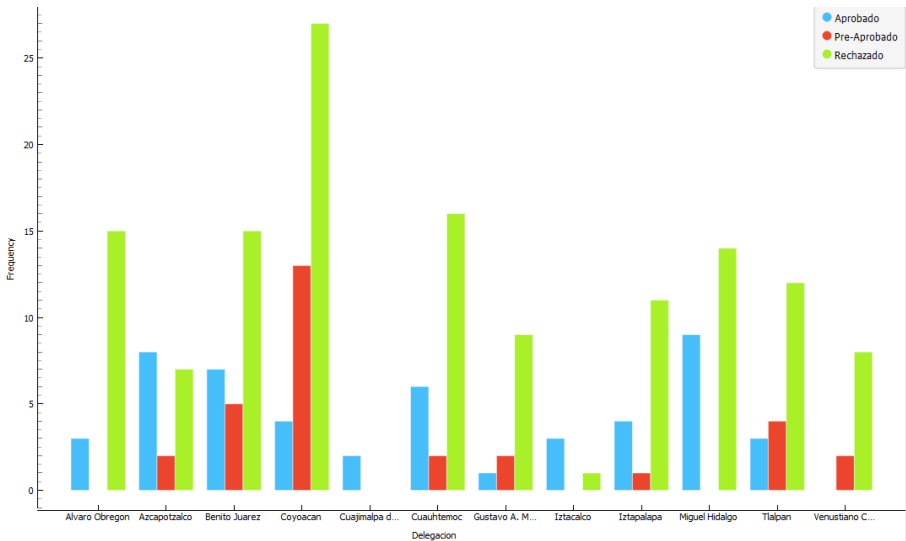


Figura 4: Gráfica de estatus final de solicitudes de crédito por alcaldía.

Si ahora tomamos en cuenta el tipo de vivienda en la que los clientes que realizaron las solicitudes de crédito, ya sea vivienda familiar o vivienda propia en la que se tenga o no una cuota periódica como renta o hipoteca, podemos visualizar que:

La delegación Venustiano Carranza presenta una mayor tasa de personas (70%) que están en viviendas donde no tienen una cuota periódica y las delegaciones Benito Juárez, Coyoacán, Cuajimalpa de Morelos, Iztacalco, Iztapalapa, Miguel Hidalgo y Tlalpan presentan una división muy similar en cuanto al tipo de vivienda, pues la mitad o más de la mitad tienen vivienda familiar o propia.

Por el contrario en la delegación Álvaro Obregón hay un mayor índice de personas, tiene que pagar mensualmente el lugar donde viven.

Delegacion	Familiares	Hipoteca	Propia	Renta	Total general	Tasa vivienda s/p
Alvaro Obregon	5	5	3	5	18	44.44444444
Azcapotzalco	5	5	3	4	17	47.05882353
Benito Juarez	1	6	13	7	27	51.85185185
Coyoacan	12	13	12	7	44	54.54545455
Cuajimalpa de Morelos	1	1			2	50
Cuauhtemoc	8	5	6	5	24	58.33333333
Gustavo A. Madero	1	3	2	6	12	25
Iztacalco	1		1	2	4	50
Iztapalapa	3	4	5	4	16	50
Miguel Hidalgo	8	6	4	5	23	52.17391304
Tlalpan	5	2	5	7	19	52.63157895
Venustiano Carranza	7			3	10	70
Total general	57	50	54	55	216	51.38888889

Figura 5: Tabla de viviendas de clientes por alcaldía.

Podemos observar que los resultados obtenidos pueden ser verse afectados según la alcaldía en la cual se esté realizando el análisis de la información, para poder observar de mejor manera tenemos la siguiente gráfica:

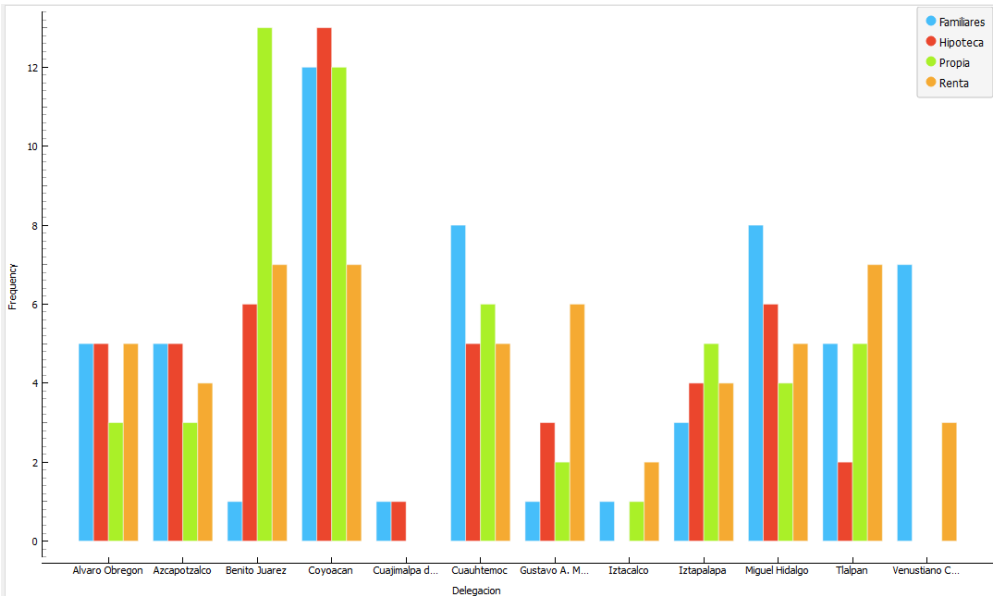


Figura 6: Gráfica de viviendas de clientes por alcaldía.

Prosiguiendo el análisis esta vez con el grado de escolaridad de los clientes podemos observar que dependiendo de la delegación se tiene un grado de escolaridad, por ejemplo la delegación Coyoacán presenta un mayor número de clientes con educación superior, 22. Mientras que Benito Juárez tiene 19 y Tlalpan 16.

Escolaridad	Licenciatura	Posgrad	Preparatori	Secundari	Total genera	Tasa estudios sup
Alvaro Obregon	5	6	6	1	18	61.11111111
Azcapotzalco	5	7	5		17	70.58823529
Benito Juarez	8	11	8		27	70.37037037
Coyoacan	12	10	19	3	44	50
Cuajimalpa de Morelos		2			2	100
Cuauhtemoc	5	10	9		24	62.5
Gustavo A. Madero	4	5	3		12	75
Iztacalco	1	2	1		4	75
Iztapalapa	5	6	5		16	68.75
Miguel Hidalgo	5	4	13	1	23	39.13043478
Tlalpan	9	7	3		19	84.21052632
Venustiano Carranza	1	5	4		10	60
Total general	60	75	76	5	216	62.5

Figura 7: Tabla de grado de educación de clientes por alcaldía.

Pero si realizamos un análisis con la tasa de personas con educación superior encontramos que en la delegación Cuajimalpa de Morelos, ambos clientes tienen una educación de posgrado mientras que en Tlalpan de 10 clientes 8 tienen educación superior, siendo estas dos delegaciones las que tienen una mejor tasa de personas con educación superior.

A pesar de que Coyoacán presenta un mayor número de personas con educación superior es la delegación que tiene menor tasa de este índice al contar con un 50%.

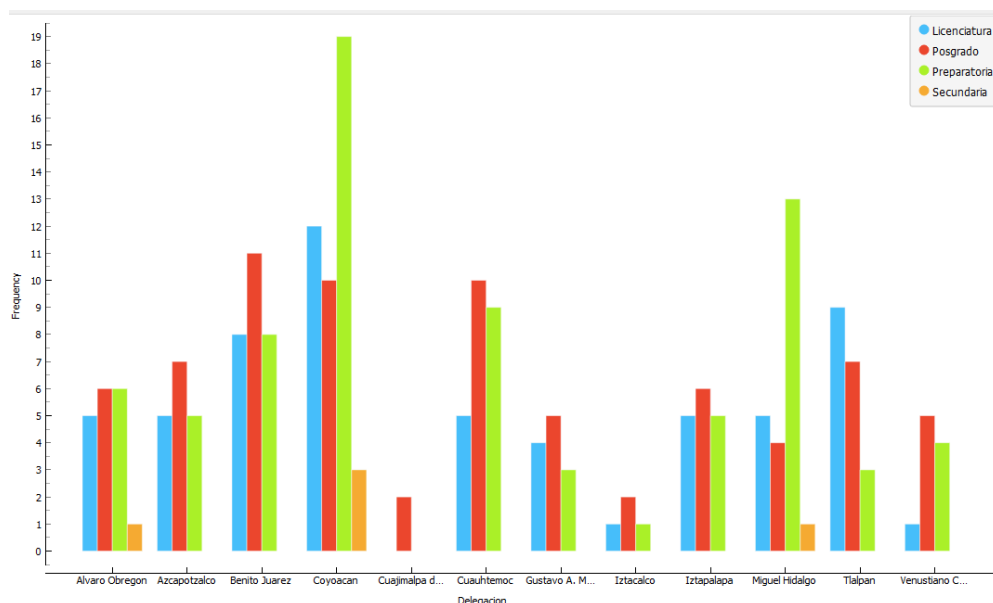


Figura 8: Gráfica de grado de escolaridad de clientes por alcaldía.

Ahora si realizamos un promedio de ingresos inferidos tenemos que el promedio general es \$45,195.54167 pesos mexicanos.

En cuanto al promedio por delegaciones: Iztacalco cuenta con el menores ingresos, al tener solo \$20,956 pesos mexicanos, mientras que la alcaldía Gustavo A. Madero es quien más sobrepasa el promedio contando con una cifra de \$51,713.08 pesos mexicanos.

Etiquetas de fila	Promedio de INGRESO_INFERIDO
Alvaro Obregon	34701.11111
Azcapotzalco	47312.35294
Benito Juarez	41036
Coyoacan	47349.88636
Cuajimalpa de Morelos	56038
Cuauhtemoc	48139.5
Gustavo A. Madero	51713.08333
Iztacalco	20956
Iztapalapa	43800
Miguel Hidalgo	50142.65217
Tlalpan	50169.10526
Venustiano Carranza	36284.1
Total general	45195.54167

Figura 9: Tabla de ingresos inferidos de clientes por alcaldía.

Realizando un análisis con la distribución normal y el diagrama de caja de gato, encontramos que la delegación que presenta un mayor rango de ingresos inferidos para sus clientes es Azcapotzalco, con ingresos centrales desde los \$18,452 hasta los \$74,287 pesos mexicanos.

La delegación Iztacalco presenta un sesgo hacia la izquierda al tener tanto clientes que rondan el ingreso de \$20,000. Mientras que la alcaldía Gustavo A. Madero presenta un ligero sesgo hacia la derecha.

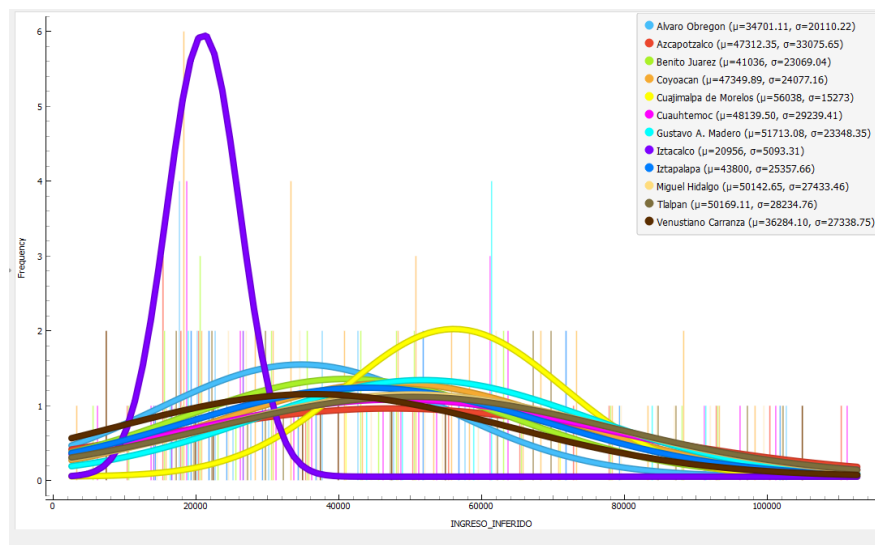


Figura 10: Distribución normal.

Las delegaciones que presentan una mejor consistencia en la distribución del ingreso inferido de sus clientes son Cuajimalpa de Morelos, Azcapotzalco, Coyoacán, Tlalpan y Miguel Hidalgo.

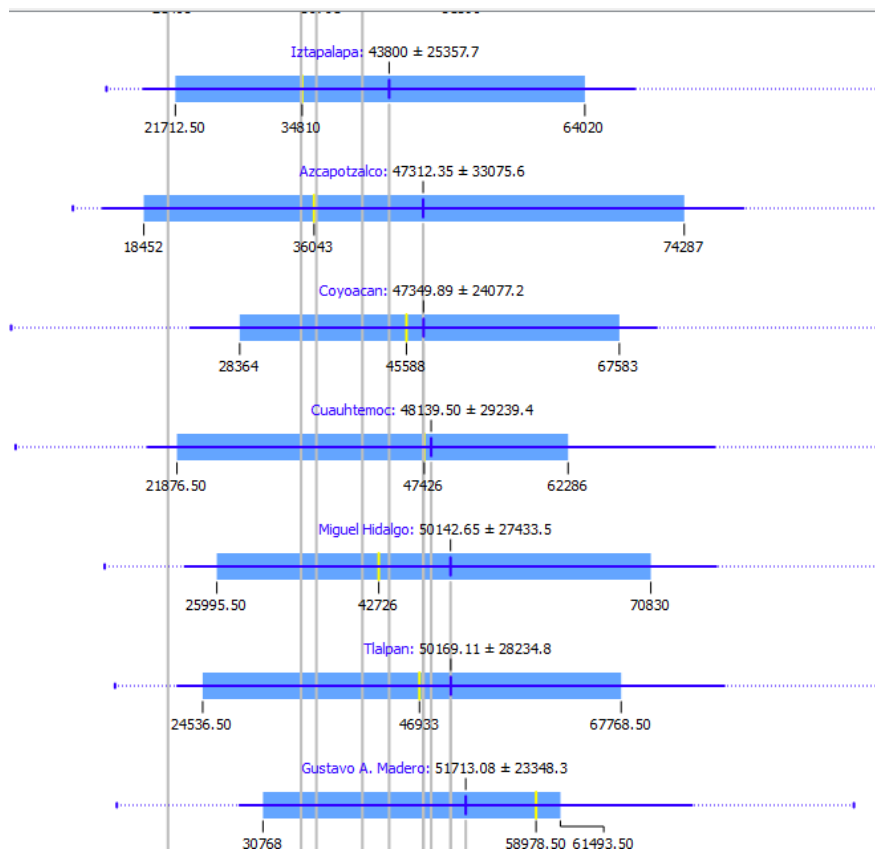


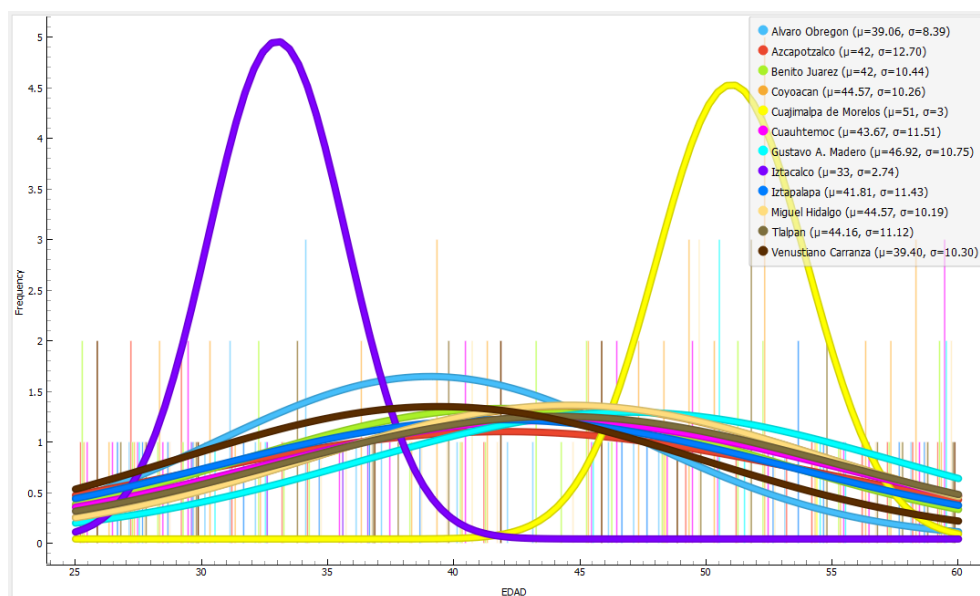
Figura 11: Diagrama de caja de gato.

Hacemos un análisis de las edades de los clientes, realizando una tabla de distribución, pero primero analizando los datos podemos ver que en la alcaldía Cuajimalpa de Morelos presenta un promedio de edad de 51 años en sus clientes, siendo la alcaldía con más clientes longevos, seguida de Gustavo A. Madero con 47 años aproximadamente. La alcaldía que cuenta con un promedio de clientes más jóvenes es Iztacalco, con 33 años. Finalmente, podemos observar que el promedio general de los clientes es de 43 años aproximadamente.

Etiquetas de fila	Promedio de EDAD
Alvaro Obregon	39.05555556
Azcapotzalco	42
Benito Juarez	42
Coyoacan	44.56818182
Cuajimalpa de Morelos	51
Cuauhtemoc	43.66666667
Gustavo A. Madero	46.91666667
Iztacalco	33
Iztapalapa	41.8125
Miguel Hidalgo	44.56521739
Tlalpan	44.15789474
Venustiano Carranza	39.4
Total general	42.98148148

Figura 12: Tabla de edades de acuerdo a los clientes por alcaldías.

En la tabla de distribución se observa como la delegación Iztacalco se encuentra sesgada hacia la izquierda, mientras que la delegación Cuajimalpa de Morelos y Gustavo A. Madero se encuentran sesgados hacia la derecha esto en consecuencia de ser la alcaldía con más clientes longevos



Con estos datos obtenidos podemos hacer un análisis más profundo aplicando modelos más específicos para poder llegar a un resultado aproximado o conciso para saber la solución a la problemática inicial.

Aplicación de modelos

- *Árbol de decisión*

Comenzamos estableciendo una relación entre los clientes, su alcaldía, el tipo de vivienda, grado de estudios junto con la tasa de aprobación o rechazo de su solicitud de crédito en el banco. Una vez con esos datos podemos realizar un árbol de decisión.

Dicho árbol se realizará en el software Orange, el cuál se utilizó a lo largo de todo el curso, por lo que de manera sencilla agregamos los datos y realizamos el árbol.

Dando como resultado un árbol bastante extenso donde la primera segmentación se divide de un lado las alcaldías Benito Juárez, Gustavo A. Madero, Coyoacán, Tlalpan o Venustiano Carranza, en donde se aprecian los siguientes datos (Figura 14).

Análisis de resultados

En general, dentro de estas alcaldías existe una tasa de rechazos general del 63.4%.

Sin importar la alcaldía, el crédito es rechazado cuando se tiene un nivel de estudios de secundaria. Pero si se tiene un grado de posgrado o secundaria, hay una probabilidad del 75.5% de ser rechazados en el crédito, mientras que si se tiene un grado de escolaridad de preparatoria o licenciatura, la probabilidad se reduce a 56.3%.

1. De estas alcaldías, Benito Juárez es la que presenta una mayor tasa de aceptación para personas con preparatoria o licenciatura al tener 56.2%. Cabe destacar que si una persona con licenciatura dentro de la misma alcaldía renta su vivienda, la solicitud del crédito se le rechaza, pero si tiene una vivienda propia o con hipoteca hay un 50% de aprobación del crédito.

En Coyoacán para personas con una vivienda propia o familiar hay una tasa de aprobación del 45%, mientras que para aquellas personas que están rentando muestran una tasa de pre aprobación del 50% y una de aprobación del 25%.

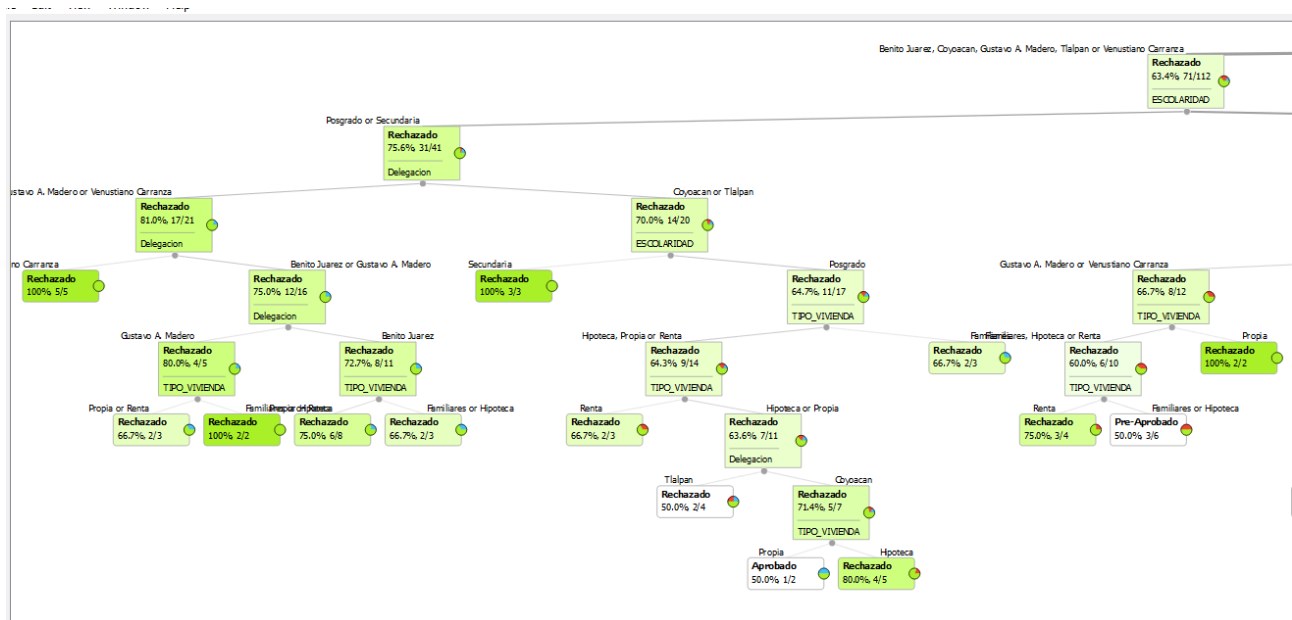


Figura 14: Árbol de decisión.

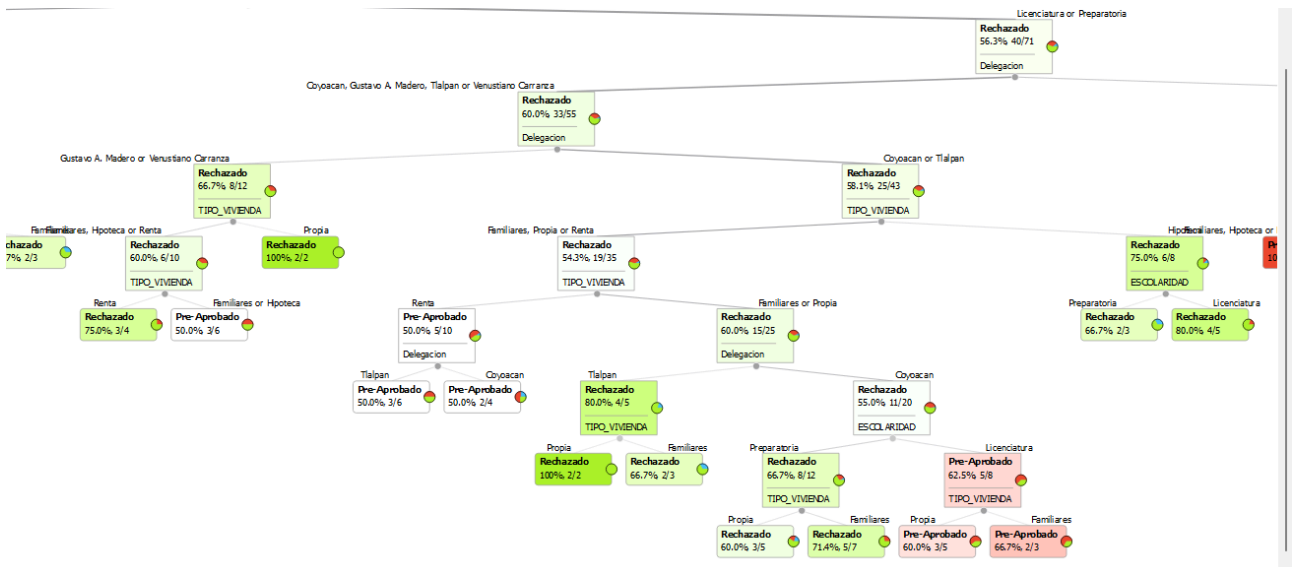


Figura 15: Árbol de decisión.

2. También podemos observar que por otro lado, las alcaldías Álvaro Obregón, Azcapotzalco, Cuajimalpa de Morelos, Cuauhtémoc, Iztacalco, Iztapalapa y Miguel Hidalgo, la tasa de aceptación es del 38.5%, 1.9% mayor en comparación a las delegaciones del primer punto. En estas alcaldías se presentan los siguientes datos:

- En la delegación Iztapalapa, aquellas personas que rentan su vivienda tienen un 75% de probabilidad de ser aceptados en su crédito, mientras que en la delegación Cuauhtémoc con las mismas condiciones, la probabilidad es del 60%.
- En la delegación Miguel Hidalgo aquellas personas con un grado de estudios de posgrado y con una hipoteca tienen una probabilidad de aprobación del 75%.
- En la delegación Álvaro Obregón aquellas personas que cuenten con una licenciatura y cuenten con una casa familiar o Hipotecada su crédito es aceptado.
- La delegación Azcapotzalco cuenta con un 50% de aprobación para personas con vivienda propia o con hipoteca.

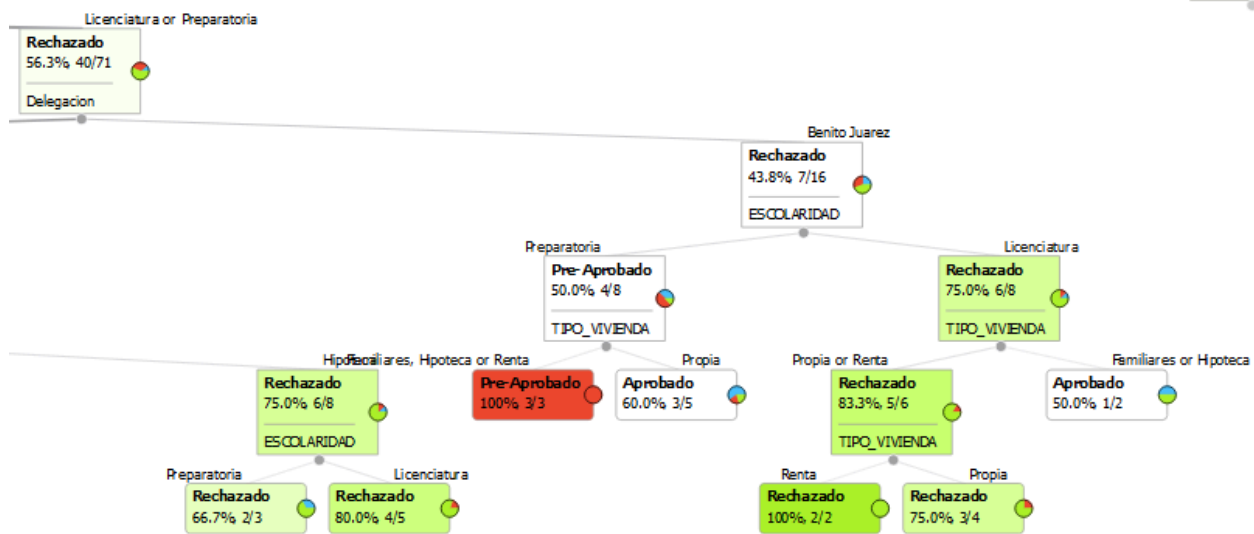


Figura 17: Árbol de decisión.

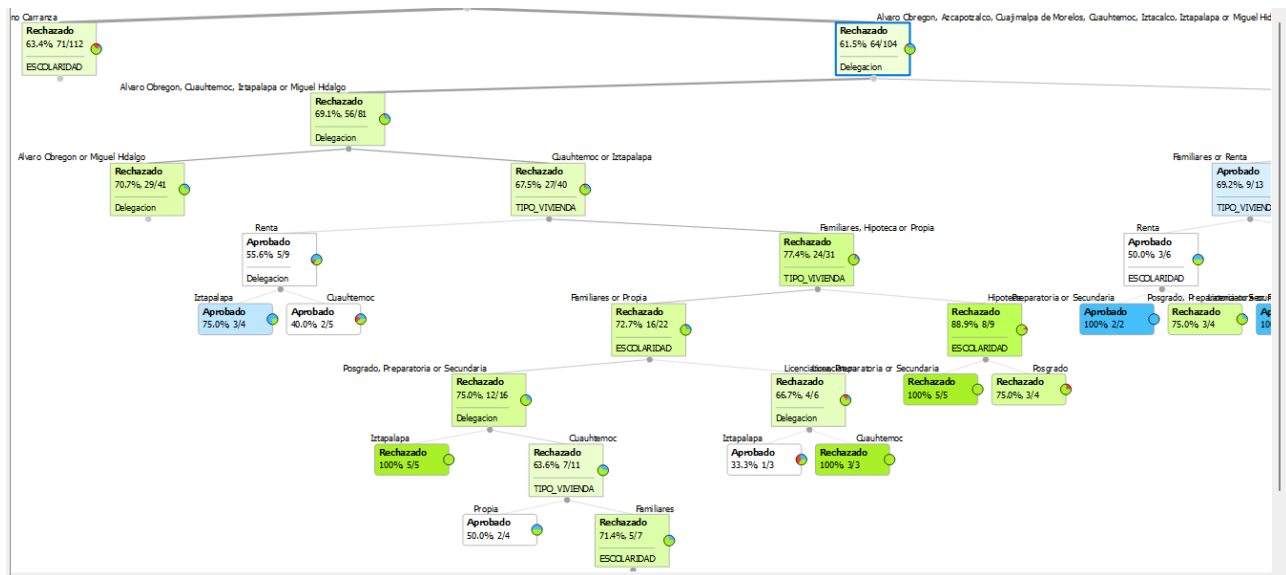


Figura 18: Árbol de decisión.

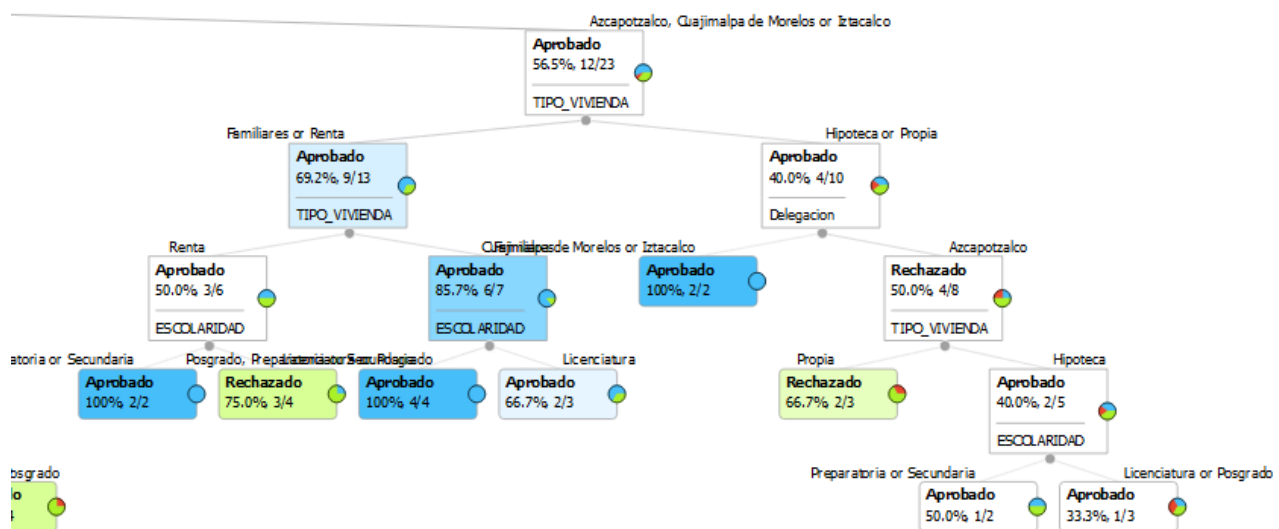


Figura 19: Árbol de decisión.

Un factor importante que se debe tomar en cuenta es que todos aquellos clientes catalogados con un Score de cliente menor o igual a 101 fueron rechazados. Mientras que a solo 1 persona con un score de cliente mayor 101 se le rechazó el crédito, de esta forma se demuestra que el score es fundamental a la hora de seleccionar si a un cliente se le rechaza o se le aprueba su crédito.

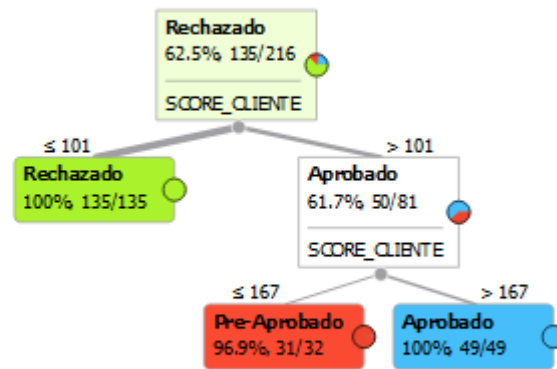


Figura 20: Árbol de decisión.

- SVM

Para determinar si hay alcaldías en las cuales haya un mayor ingreso inferido y haya una mayor capacidad de pago, se utilizará el modelo SVM, por lo que podemos determinar lo siguiente:

- La mayor parte de las personas en las alcaldías sin importar sus ingresos inferidos tienen una capacidad de pago pequeña.
- La delegación Iztacalco, presenta una pendiente negativa abrupta, quiere decir que a menor ingreso inferido existe una capacidad de pago mayor, quedando como límite un ingreso inferido de \$4,000 pesos mexicanos. Otras delegaciones que presentan una pendiente negativa son las alcaldías Benito Juárez, Coyoacán y Miguel Hidalgo.
- Por el lado contrario la delegación Álvaro Obregón y Azcapotzalco presentan una pendiente positiva lo que nos dice que la gente de estas zonas a mayor ingreso inferido existe una mayor capacidad de pago, el cliente con mayor ingreso inferido es de \$100,600 con una capacidad de pago de 1.20 y una edad de 58 años. Clasificado como un cliente regular, esto puede ser consecuencia de su edad y el tipo de vivienda que no se muestra cuál es.
- Iztapalapa y Cuauhtémoc también tienen una pendiente positiva, aunque está por debajo de las dos alcaldías antes mencionadas.

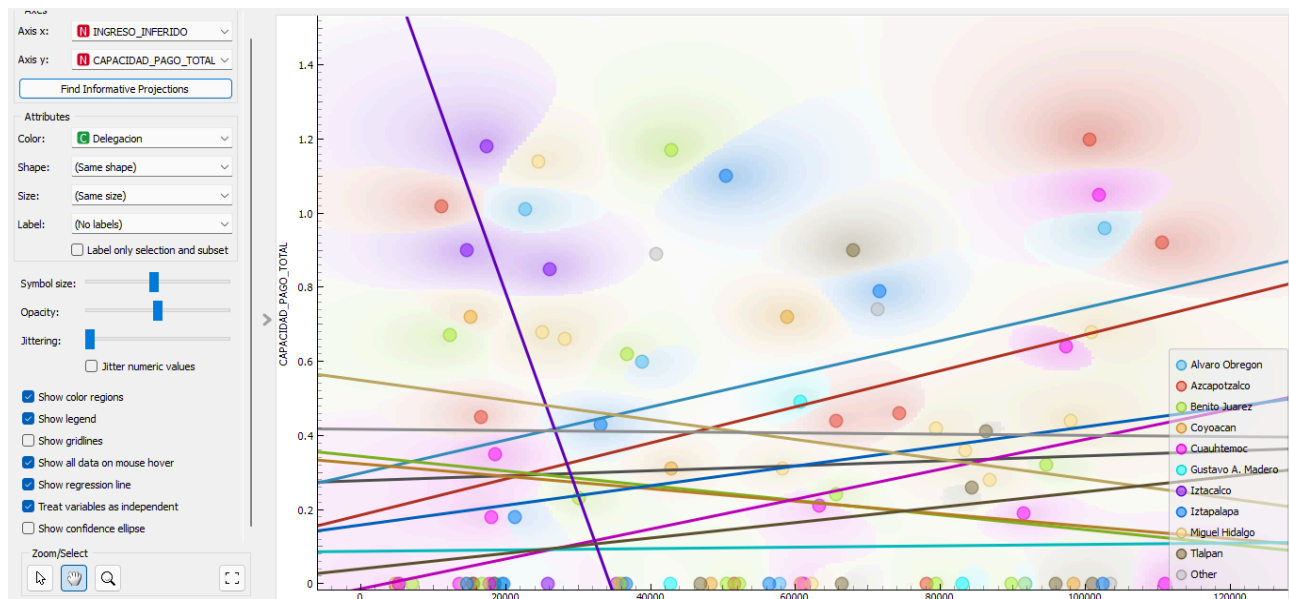


Figura 21: SVM.

Analizamos el comportamiento de los clientes, de lo que pueden pagar con sus posibles deudas que puedan adquirir. Para ello, comparamos la suma de líneas resolventes con respecto de la capacidad de pago total. Arrojando los siguientes resultados:

- Las alcaldías Miguel Hidalgo y Azcapotzalco tienen un gran poder adquisitivo al tener clientes con grandes cantidades de suma líneas revolventes; sin embargo, las sucursales de la delegación Miguel Hidalgo corren grandes riesgos, ya que sus clientes tienen una capacidad de pago por debajo del 0.7. Mientras que en la delegación Azcapotzalco a pesar de haber clientes con una gran cantidad de suma líneas resolventes estos tienen una alta capacidad de pago lo cual habla muy bien del perfil financiero de la zona, teniendo personas con alto poder adquisitivo y una oportunidad de crecimiento en cuanto a créditos revolventes.
- También existen múltiples alcaldías en donde sus clientes sin importar el nivel de capacidad de pago total estos tienen líneas de crédito revolventes lo que nos dice que el cliente no tiene líneas de crédito activas que se puedan reutilizar como lo son tarjetas de crédito y el crédito que se le otorgó es de plazo fijo (préstamo personal, de un auto o hipoteca).

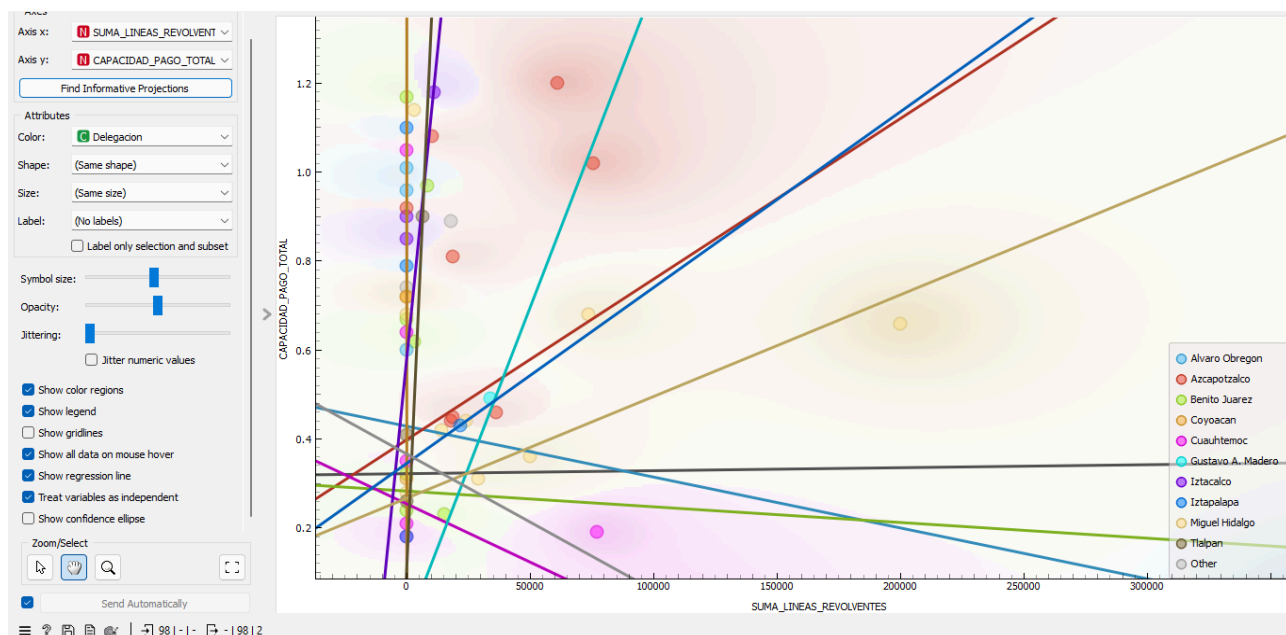


Figura 22: SVM.

Por último, estableciendo la relación de la línea de crédito final otorgada y la capacidad de pago, tenemos comportamientos similares a los ya establecidos anteriormente:

- Álvaro Obregón y Azcapotzalco tienen clientes con una alta línea de crédito final, pero también es proporcional con su capacidad de pago.
- Alcaldías como Cuauhtémoc, Benito Juárez e Iztacalco tienen una baja línea de crédito final, pero tienen una alta capacidad de pago.

● Clustering

Para realizar el clustering, se dividió en 3 grupos con las instancias de edad, delegación, score cliente, tipo vivienda y escolaridad, encontramos los siguientes comportamientos:

- El cluster C1 agrupa clientes con scores de cliente más bajos, en donde los clientes con un mayor nivel de score tiene una menor línea de crédito final, lo cual puede representar clientes con mayor riesgo o menor estabilidad financiera. Aquí se encuentran en su mayoría clientes de la alcaldía Coyoacán y Gustavo A. Madero.
- El cluster C2 representa a clientes con scores más altos y con ingresos inferidos más altos, esto nos indica que son clientes con un mayor poder adquisitivo teniendo que ser clientes con un menor riesgo financiero. En este cluster las delegaciones que predominan son Miguel Hidalgo y Azcapotzalco.

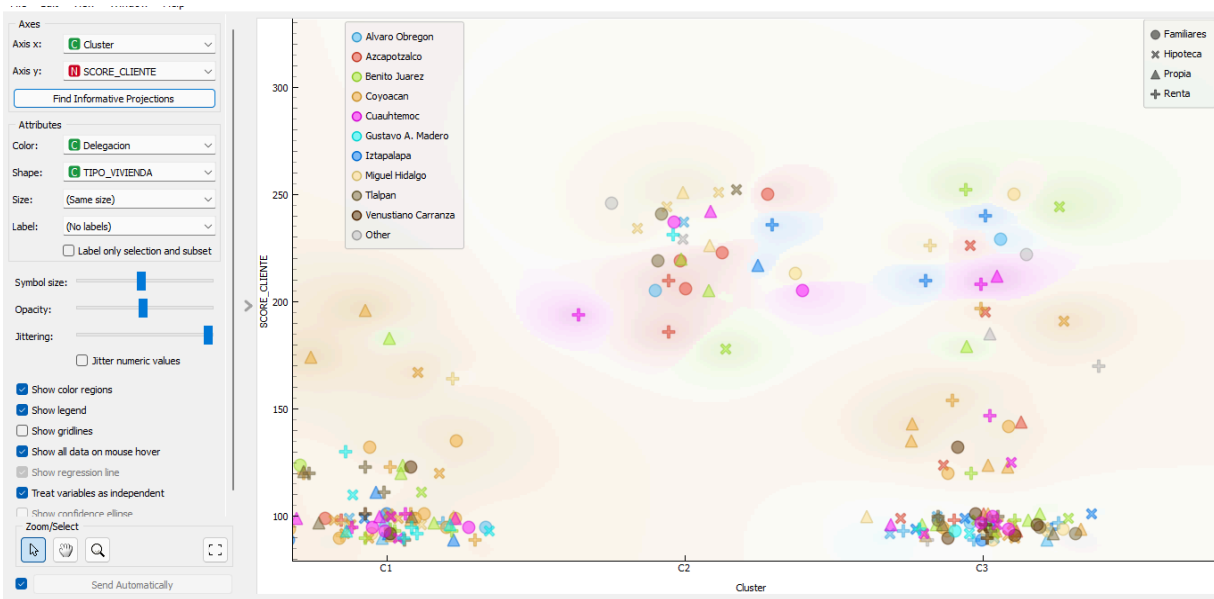


Figura 23: Clustering.

- Modelos Lineales

Para establecer una relación entre la calificación del cliente y la capacidad de pago total realizamos un modelo lineal. Para ello nos apoyamos de Polynomial Regression seleccionando el grado del polinomio que más se ajusta , en este caso es de grado 5.

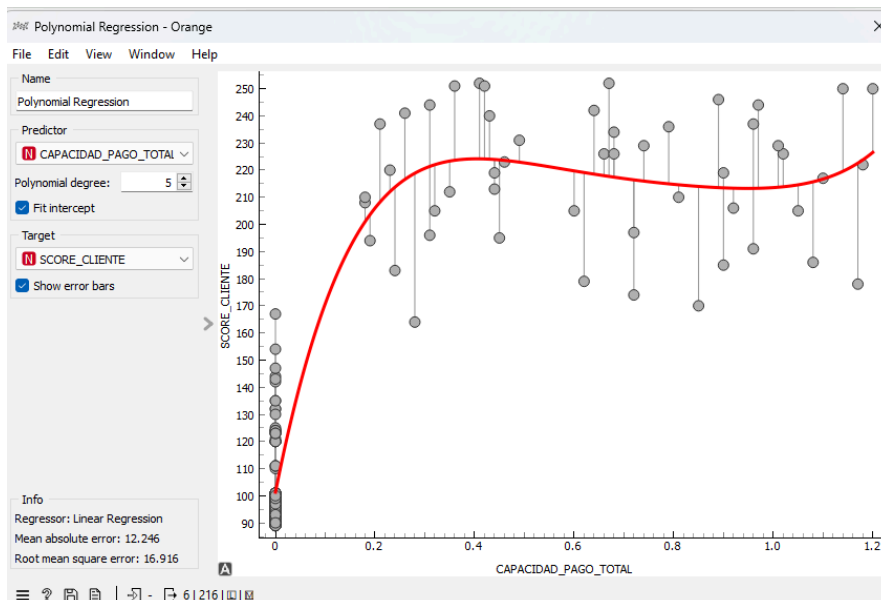


Figura 24: Modelo lineal.

Gracias al widget data table observamos que la función polinomial que describe la recta es:

$$\text{scoreCliente} = 922.919\text{capacidadPagoTot} - 2611.58(\text{capacidadPagoTot})^2 + 3517.2(\text{capacidadPagoTot})^3 - 2342.19(\text{capacidadPagoTot})^4 + 625.909(\text{capacidadPagoTot})^5 + 101.3$$

	name	coef
1	intercept	101.3
2	CAPACIDAD_P...	922.919
3	CAPACIDAD_P...	-2611.58
4	CAPACIDAD_P...	3517.2
5	CAPACIDAD_P...	-2342.19
6	CAPACIDAD_P...	625.909

Figura 25: Función polinomial.

Donde:

- Si el resultado de la función anteriormente mencionada es 0 describe a la función polinomial.
- Si el resultado es <0 el punto cae por encima de la función lineal lo que nos indica que el score de cliente es suficientemente bueno para su capacidad de pago total, teniendo a clientes con riesgo bajo.
- Si el resultado es >0 el punto cae por debajo de la función lineal esto puede ser de clientes que a pesar de tener un gran saldo inferido y poca suma en líneas revolventes existen otros factores como la edad o tipo de vivienda para concluir que un cliente tiene menor calificación.

Con esta gráfica podemos concluir que los datos al encontrarse dispersos no existe una estrecha relación entre la capacidad de pago total y la calificación del cliente, sin embargo, esta dispersión no es muy ajena a la función establecida por lo que para determinar la calificación del mismo influyen otros factores.

En vista de que no solo podemos considerar una sola variable para determinar la calificación del cliente, recurrimos a una regresión lineal, dicha que involucre la edad, el ingreso inferido, la línea de crédito final, la suma de líneas revolventes y la capacidad de pago total.

El resultado de la regresión lineal con las instancias mencionadas es la siguiente:

$$\text{scoreCliente} = -0.376629\text{Edad} + 0.0002872\text{IngresoInferido} + 0.0002861\text{LineaCF}$$

$$+ 3.15669 \text{SumaLineasRevolventes} + 122.356 \text{CapacidadPagoTot} + 110.336$$

	name	coef
1	intercept	110.336
2	EDAD	-0.376629
3	INGRESO_INFERIDO	0.000287235
4	LINEA_CREDITO_FINAL	0.000286139
5	SUMA_LINEAS_REVOLVENTES	3.15669e-05
6	CAPACIDAD_PAGO_TOTAL	122.356

Figura 26: Datos para la regresión lineal.

En esta función podemos observar que para determinar la calificación del cliente la instancia que mayor peso tiene es capacidad de pago total; sin embargo, y como lo dedujimos, existen otros pesos significativos tales como suma de líneas, revolventes y edad.

- Reglas de asociación

Al querer establecer reglas de asociación entre las instancias de alcaldía, la escolaridad, el tipo de vivienda y aprobación de la tarjeta de crédito encontramos los siguientes resultados:

- Con una confianza del 100% y un soporte del 2.3% , como lo habíamos visto en el árbol de decisión, si una persona tiene solo la secundaria, la tarjeta será rechazada.
- Para la alcaldía Cuauhtémoc, con un soporte del 2.3% podemos decir que toda persona con un grado de licenciatura ha sido rechazada. Caso similar sucede con la alcaldía Álvaro Obregón, solo que esta vez para personas con posgrado.
- Continuando con la alcaldía Álvaro Obregón con un soporte del 2.3% todas las personas que rentan fueron rechazadas para el trámite.
- En la alcaldía Iztapalapa aquellas personas que tienen un grado de estudios de preparatoria fueron rechazadas.
- Mientras que en la alcaldía Azcapotzalco, con un soporte del 1.9% un 80% de las personas que tienen una vivienda familiar son aceptadas para la tarjeta de crédito, mientras que para las personas con una vivienda propia estas tienen una tasa de rechazo de del 66.7%.

Info						Antecedent		Consequent
Rules: 3574 (shown 1232)						Delegacion=Alvaro Obregon, ESCOLARIDAD=Posgrado		→ APROBACION_TC=Rechazado
Find association rules						Delegacion=Cuauhtemoc, ESCOLARIDAD=Licenciatura		→ APROBACION_TC=Rechazado
Min. supp.: 0.0001 %						Delegacion=Alvaro Obregon, TIPO_VIVIENDA=Renta		→ APROBACION_TC=Rechazado
Min. conf.: 1 %						Delegacion=Iztapalapa, ESCOLARIDAD=Preparatoria		→ APROBACION_TC=Rechazado
Max. rules: 100k						Delegacion=Venustiano Carranza, ESCOLARIDAD=Posgrado		→ APROBACION_TC=Rechazado
<input type="checkbox"/> Induce only classification rules						ESCOLARIDAD=Secundaria		→ APROBACION_TC=Rechazado
<input checked="" type="checkbox"/> Restrict search by below filters						Delegacion=Iztapalapa, TIPO_VIVIENDA=Hipoteca		→ APROBACION_TC=Rechazado
Find Rules						Delegacion=Gustavo A. Madero, ESCOLARIDAD=Licenciatura		→ APROBACION_TC=Rechazado
Filter by Antecedent						Delegacion=Alvaro Obregon, ESCOLARIDAD=Posgrado, TIPO_VIVIENDA=Familiares		→ APROBACION_TC=Rechazado
Contains:						Delegacion=Alvaro Obregon, TIPO_VIVIENDA=Propia		→ APROBACION_TC=Rechazado
Items, min: 1 max: 998						Delegacion=Iztapalapa, TIPO_VIVIENDA=Familiares		→ APROBACION_TC=Rechazado
Filter by Consequent						Delegacion=Iztapalapa, ESCOLARIDAD=Preparatoria, TIPO_VIVIENDA=Hipoteca		→ APROBACION_TC=Rechazado
Contains: aprobacion_tc						Delegacion=Venustiano Carranza, ESCOLARIDAD=Posgrado, TIPO_VIVIENDA=Familiares		→ APROBACION_TC=Rechazado
Items, min: 1 max: 2						Delegacion=Venustiano Carranza, TIPO_VIVIENDA=Renta		→ APROBACION_TC=Rechazado
Send selection						Delegacion=Coyoacan, ESCOLARIDAD=Secundaria		→ APROBACION_TC=Rechazado
0.028	1.000	0.028	22.500	1.600	0.010	ESCOLARIDAD=Secundaria, TIPO_VIVIENDA=Familiares		→ APROBACION_TC=Aprobado
0.023	1.000	0.023	27.000	1.600	0.009	Delegacion=Azcapotzalco, ESCOLARIDAD=Posgrado, TIPO_VIVIENDA=Familiares		→ APROBACION_TC=Rechazado
0.023	1.000	0.023	27.000	1.600	0.009	Delegacion=Benito Juarez, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Renta		→ APROBACION_TC=Rechazado
0.023	1.000	0.023	27.000	1.600	0.009	Delegacion=Cuauhtemoc, ESCOLARIDAD=Posgrado, TIPO_VIVIENDA=Familiares		→ APROBACION_TC=Rechazado
0.023	1.000	0.023	27.000	1.600	0.009	Delegacion=Cuauhtemoc, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Hipoteca		→ APROBACION_TC=Rechazado
0.023	1.000	0.023	27.000	1.600	0.009	Delegacion=Cuauhtemoc, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Propia		→ APROBACION_TC=Rechazado
0.019	1.000	0.019	33.750	1.600	0.007	Delegacion=Miguel Hidalgo, ESCOLARIDAD=Preparatoria, TIPO_VIVIENDA=Hipoteca		→ APROBACION_TC=Aprobado
0.019	1.000	0.019	33.750	1.600	0.007	Delegacion=Miguel Hidalgo, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Familiares		→ APROBACION_TC=Rechazado
0.014	1.000	0.014	45.000	1.600	0.005	Delegacion=Miguel Hidalgo, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Hipoteca		→ APROBACION_TC=Rechazado
0.014	1.000	0.014	45.000	1.600	0.005	Delegacion=Miguel Hidalgo, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Propia		→ APROBACION_TC=Rechazado
0.014	1.000	0.014	45.000	1.600	0.005	Delegacion=Tlalpan, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Propia		→ APROBACION_TC=Rechazado
0.014	1.000	0.014	45.000	1.600	0.005	Delegacion=Alvaro Obregon, ESCOLARIDAD=Posgrado, TIPO_VIVIENDA=Hipoteca		→ APROBACION_TC=Rechazado
0.014	1.000	0.014	45.000	1.600	0.005	Delegacion=Alvaro Obregon, ESCOLARIDAD=Preparatoria, TIPO_VIVIENDA=Hipoteca		→ APROBACION_TC=Rechazado
0.009	1.000	0.009	25.000	4.320	0.007	Delegacion=Alvaro Obregon, ESCOLARIDAD=Licenciatura, TIPO_VIVIENDA=Renta		→ APROBACION_TC=Rechazado
0.009	1.000	0.009	67.500	1.600	0.003			

Figura 27: Reglas de asociación.

Con los resultados obtenidos podemos realizar un business case, el cual nos va a permitir plantear de mejor manera la solución llegando a conclusiones de cada caso analizado.

Conclusiones

Podemos concluir que se realizó la planeación de proyecto de forma correcta, en la que además de desglosar el problema logramos analizar la información para llegar a la solución propuesta logramos tener una buena comunicación como equipo para poder desarrollar el proyecto satisfactoriamente así cumpliendo los objetivos propuestos en la introducción.

Logramos aplicar los conocimientos adquiridos durante todo el curso de Minería de Datos lo que nos permitió hacer un análisis de los datos de manera que nos permitió proponer una solución con el fin de poder conocer el comportamiento financiero y social en las alcaldías de la Ciudad de México utilizando herramientas como el software Orange y Excel.