Tecnológico de Monterrey

05 de mayo de 2024

Instituto Tecnológico y de Estudios Superiores de Monterrey

Maestría en Inteligencia Artificial Aplicada – MNA

Materia: Proyecto Integrador

Avance 1. Análisis exploratorio de datos

Equipo:

Ruiz González Rodrigo - A01793081

Daniel Hernández Mora - A01793538

Juan Sebastián Téllez López - A01793859

Profesora Titular:

Dra. Grettel Barceló Alonso

EDA - Exploratory Data Analysis

El análisis de datos exploratorios (EDA) es un paso inicial crucial en cualquier proyecto de análisis de datos. Implica un conjunto de técnicas para familiarizarse con sus datos, descubrir sus características e identificar patrones o tendencias potenciales (Martínez, 2005).

EDA tiene que ver con la exploración y la investigación más que con la prueba o confirmación de hipótesis:

1. **Comprensión de la estructura de datos:** EDA le ayuda a comprender la estructura general de su conjunto de datos. Esto incluye identificar los tipos de datos de cada variable, observar la presencia de valores faltantes o valores atípicos y explorar la distribución de valores dentro de cada variable.
2. **Descubrir relaciones:** EDA le permite descubrir relaciones entre diferentes variables en su conjunto de datos. Esto podría implicar visualizar relaciones a través de diagramas de dispersión o matrices de correlación, que pueden ayudarlo a identificar posibles correlaciones o dependencias entre variables.
3. **Detección de anomalías:** EDA es útil para detectar anomalías dentro de su conjunto de datos. Estas anomalías podrían ser valores atípicos, puntos de datos faltantes o patrones inesperados. Identificar estas anomalías le ayuda a decidir cómo manejarlas durante las últimas etapas de su análisis.
4. **Formulación de hipótesis:** a través de EDA, puede descubrir patrones o tendencias interesantes dentro de sus datos. Esto puede generar nuevas preguntas y guiarlo en la formulación de hipótesis que se probarán mediante métodos estadísticos más adelante en su análisis.
5. **Preparación de limpieza de datos:** EDA puede ayudar a preparar sus datos para un análisis posterior. Al identificar valores faltantes o valores atípicos, puede decidir cuáles son las técnicas de limpieza de datos adecuadas para abordar estos problemas antes de continuar con el modelado u otros métodos de análisis.

En esencia, EDA es fundamental para sentar las bases para un análisis de datos exitoso. En una comprensión más profunda de sus datos a través de EDA, puede tomar decisiones informadas sobre el manejo de los datos, seleccionar técnicas de modelado y, finalmente, sacar conclusiones significativas de su análisis (Behrens, 2003).

Beneficios de realizar EDA:

- **Decisiones informadas:** una comprensión profunda de sus datos le permite tomar decisiones informadas sobre el manejo de los datos, la selección de técnicas de modelado y, en última instancia, sacar conclusiones significativas de su análisis.

- **Identificación de problemas:** EDA le ayuda a descubrir problemas potenciales con sus datos desde el principio, ahorrándole tiempo y esfuerzo en el futuro.
- **Generación de conocimientos:** EDA puede generar nuevas ideas y llevarlo a descubrimientos inesperados dentro de sus datos.
- **Comunicación mejorada:** una comprensión clara de sus datos le permite comunicar mejor los hallazgos y los conocimientos a los demás.

Los clientes pueden obtener numerosos beneficios al incorporar el análisis de datos exploratorios (EDA) en un proyecto. EDA sirve como base crucial para obtener información significativa a partir de los datos, lo que permite a las empresas tomar decisiones informadas que impulsen el crecimiento y mejoren las experiencias de los clientes. Es una herramienta invaluable que permite a las empresas obtener conocimientos más profundos de sus datos, lo que genera una amplia gama de beneficios para sus clientes (Tukey, 2009). Al incorporar EDA en sus proyectos, las empresas pueden tomar decisiones informadas, desarrollar productos y servicios centrados en el cliente, optimizar campañas de marketing, reducir costos y mejorar su eficiencia general. A medida que las empresas dependen cada vez más de los datos para impulsar su éxito, EDA seguirá desempeñando un papel fundamental para permitirles alcanzar sus objetivos y ofrecer experiencias excepcionales a los clientes.

Un ejemplo de caso de éxito utilizando EDA es la cadena OXXO, ya que se enfrentaba a un desafío persistente: gestionar grandes cantidades de datos de inventario para garantizar que tuvieran los productos adecuados en las tiendas adecuadas y en el momento adecuado. El exceso de existencias provocó un desperdicio de espacio de almacenamiento, capital inmovilizado en productos no vendidos y rebajas para eliminar el exceso de inventario.

Los resultados fueron notables:

1. **Aumento de la rotación de inventario:** la tasa de rotación de inventario del minorista mejoró significativamente, reduciendo el espacio de almacenamiento desperdiciado y el capital invertido en productos no vendidos.
2. **Reducción de los desabastecimientos:** las situaciones de desabastecimiento disminuyeron drásticamente, lo que generó un aumento de las ventas y una mejor satisfacción del cliente.
3. **Rentabilidad mejorada:** la rentabilidad general del minorista aumentó debido a mejores prácticas de gestión de inventario y promociones específicas.

De manera general, se analizaron nuestros datos y a continuación se detallan las tendencias y resultados importantes:

- **¿Hay valores faltantes en el conjunto de datos? ¿Se pueden identificar patrones de ausencia?**

Si, existen valores nulos o vacíos, ya que el insumo origen no contiene estos datos, sin embargo, no son cruciales para nuestro análisis y pueden ser reemplazados por 0 o no ser considerados para que no resulte error algún calculo dentro del modelo.

```
De un total de 723933 registros se encuentran los siguientes valores faltantes:
```

	Total	Porcentaje
DOCUMENTO	0	0.000000
PRIORIDAD	0	0.000000
MOTIVO_NO_PAGO	0	0.000000
DESC_ACTIVIDAD	0	0.000000
GENERO	0	0.000000
EDAD	372	0.051386
RANGO_HORA	0	0.000000
SEGMENTO_CLIENTE	0	0.000000
RIESGO_ACTUAL	133182	18.397006
RECUPERACION	0	0.000000
MORA_INICIAL	0	0.000000
MORA_ACTUAL	0	0.000000
C_RIESGO	0	0.000000
ESTADO_ICS_ACTUAL	130563	18.035233
DIAS_ACTUAL	130563	18.035233
CICLO_OBLIG	0	0.000000

- **¿Cuáles son las estadísticas resumidas del conjunto de datos?**

La mora inicial promedio es de 47.6, la mora actual promedio es de 47.19 mientras que la edad promedio es de 40.3 años, en general nuestro universo a estudiar se observa estable y con parámetros aceptables.

Resumen estadistico

```
summary_statistics = df.describe()
summary_statistics.head()
```

Python

	DOCUMENTO	PRIORIDAD	EDAD	MORA_INICIAL	MORA_ACTUAL	C_RIESGO	DIAS_ACTUAL	CICLO_OBLIG	CED_S
count	7.239330e+05	723933.000000	723561.000000	723933.000000	723933.000000	723933.000000	593370.000000	723933.000000	7.2393
mean	5.596959e+09	9.434281	40.341291	47.622902	47.195735	365.750753	45.502976	23.821544	5.5571
std	5.465188e+09	10.515613	13.474495	22.461669	33.453026	332.281866	26.738788	9.049222	5.1195
min	1.338300e+04	0.000000	-6.000000	30.000000	0.000000	-99.000000	1.000000	0.000000	3.3830
25%	1.430310e+08	3.000000	29.000000	30.000000	30.000000	0.000000	26.000000	16.000000	4.3030

- **¿Hay valores atípicos en el conjunto de datos?**

Se presentan algunos, los cuales son detallados en profundidad mas adelante, y se deja abierta la posibilidad desde la ingeniería de características, realizar agrupaciones, validaciones y garantizar que estos datos no nos sobredimensionen ni generen inconvenientes al entrenamiento del modelo.

- ¿Cuál es la cardinalidad de las variables categóricas?

```
...
C:\Users\Usuario\AppData\Local\Temp\ipykernel_18808\3935950795.py:5: FutureWarning: The
  categorical_cardinality = categorical_cardinality.append({'Columna': col, 'Cardinalid
C:\Users\Usuario\AppData\Local\Temp\ipykernel_18808\3935950795.py:5: FutureWarning: The
  categorical_cardinality = categorical_cardinality.append({'Columna': col, 'Cardinalid
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

...

	Columna	Cardinalidad
0	MOTIVO_NO_PAGO	124
1	DESC_ACTIVIDAD	481
2	GENERO	5
3	RANGO_HORA	2
4	SEGMENTO_CLIENTE	3

- ¿Existen distribuciones sesgadas en el conjunto de datos? ¿Necesitamos aplicar alguna transformación no lineal?

Algunas respecto a valores negativos o edades muy grandes, las cuales generan tendencias que no son del todo ciertas y deben ser tratadas mediante ingeniería de características, se mencionan detalladamente mas adelante en el documento.

- ¿Se identifican tendencias temporales? (En caso de que el conjunto incluya una dimensión de tiempo).

No, no estamos trabajando con medidas de tiempo para este modelo.

- ¿Hay correlación entre las variables dependientes e independientes?

Si, ya que es un dataframe de clientes, carteras en mora y gestiones de cobro.

- ¿Cómo se distribuyen los datos en función de diferentes categorías?

	Columna	Skewness
0	DOCUMENTO	0.002484
1	PRIORIDAD	1.719999
2	EDAD	NaN
3	MORA_INICIAL	1.143838
4	MORA_ACTUAL	0.416355

- ¿Existen patrones o agrupaciones (clusters) en los datos con características similares?

Algunos como para la actividad económica de los trabajadores, clasificaciones de riesgo generales o tipo de cartera del cliente.

- ¿Se deberían normalizar las imágenes para visualizarlas mejor?

No aplica.

- ¿Hay desequilibrio en las clases de la variable objetivo?

Desequilibrio en clases de variable objetivo

```
class_counts = df['RANGO_HORA'].value_counts()
class_imbalance = pd.DataFrame({'Clase': class_counts.index, 'Frecuencia': class_counts.values})
class_imbalance
```

	Clase	Frecuencia
0	Mañana	447411
1	Tarde	276522

Si se presenta un desbalanceo muy notorio, el cual tiene una razón de ser, y es que por las dinámicas del negocio y la operación la mayoría de gestiones al cliente se le realizan en la franja de la mañana, por lo tanto la tendencia va a ser a tener mas gestiones en esa franja, la cual se espera que con este modelo logre equilibrarse y de ese modo causar un impacto positivo en la operación, no solo con el mejoramiento de la tasa de éxito a que el cliente conteste, sino en la disponibilización de recursos. De igual manera para el entrenamiento del modelo se van a aplicar técnicas de balanceo de clases para garantizar muestras idénticas en ambas franjas horarias y así evitar algún tipo de sesgo con el modelo.

A continuación, se ilustrará de una forma gráfica, donde se desarrolló un tablero en Power BI **(Adjunto en la entrega)**, que nos permite observar en detalle algunas de las variables más importantes de nuestro conjunto de datos, explicando su concepto, distribución y un análisis para comprender mejor como puede aportarnos en este proyecto. Pero primero, vamos a observar de manera general la distribución de nuestro dataset:

Variables relevantes Proyecto

Prioridad

américas Business Process Services

PRIORIDAD

Tecnológico de Monterrey

¿Qué es?

Se refiere a la clasificación o nivel de importancia que se asigna a las llamadas entrantes o a los casos que se están gestionando. La priorización puede basarse en varios factores, como la urgencia del problema del cliente, el tipo de consulta o solicitud, el historial del cliente, entre otros.

TOP 5

Pioridad	Cantidad
5	252802
21	139115
20	25405
35	12945
52	9439
Total	439706

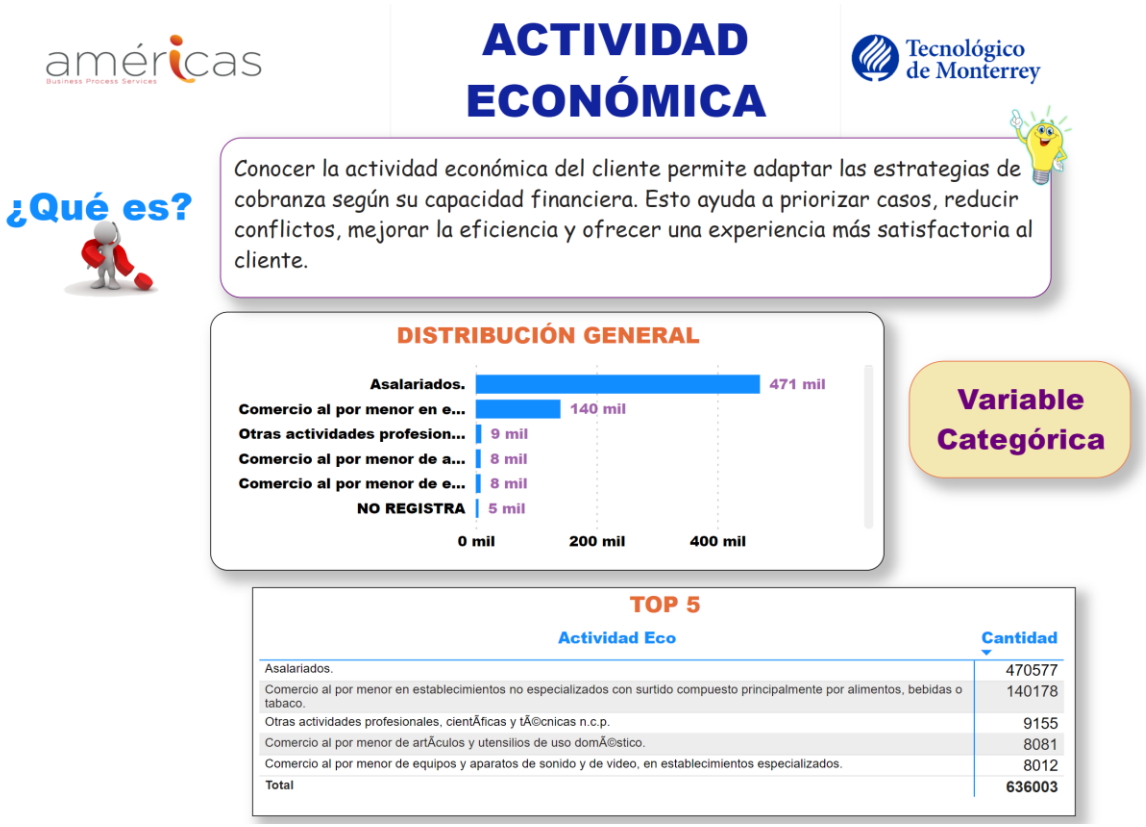
Variable Ordinal

Si bien, hay una evidente tendencia entre prioridad 5 y 21, es debido a que son clientes que están en un segmento de deuda inicial y es necesario para algunos contactarlos tan pronto como sea posible, mientras que otros no tanto, entre mayor sea el número de prioridad, mayor es la necesidad de realizar gestión sobre ellos, para garantizar una atención eficiente y satisfactoria al cliente, de momento la operación de call center utiliza este valor para ordenar los clientes y realizar la gestión. Si se llegan a presentar valores vacíos o ausencia de este número por defecto, se le asignara un 0, atendiendo a que no es tan urgente establecer contacto con él. Si se presentan valores atípicos como más de 100 o negativos, se reemplazarán a 0, ya que sería un error en la clasificación, y como se mencionó antes en la gestión, se observará una tendencia a que se marcan primero a los clientes con prioridad respecto a los demás.

Motivo de no pago

Como parte de la gestión por parte de los agentes y en un plan de que el cliente tenga una experiencia de cobro más humanizada, es muy importante entender las razones por las cuales no se ha podido efectuar el pago de las obligaciones financieras, si bien, en nuestro contexto de LATAM y específicamente en Colombia, el factor de la alta inflación, pocas oportunidades laborales o poca inteligencia financiera, ocasiona que se deba efectuar dicho proceso de cobro. Entre las principales razones por los clientes es la disminución de ingresos, pero la que se encuentra en segundo lugar “No Motivo” es muy importante mencionarla, ya que puede surgir de varios escenarios, el primero es que el cliente no manifieste una razón coherente o real del porque no ha realizado el pago, la segunda es que haya interrupciones y se corte la comunicación con el cliente antes de esa pregunta, y la tercera es cuando aún no se ha tenido gestión con el cliente a manera de abordar los valores vacíos se les asigna dicha categoría. Aunque, tenemos diversas razones que podrían convertirse en valores atípicos, se espera que en la parte de ingeniería de características se pueda agrupar razones particulares para evitar un sobredimensionamiento de esta que no aporte positivamente al modelo.

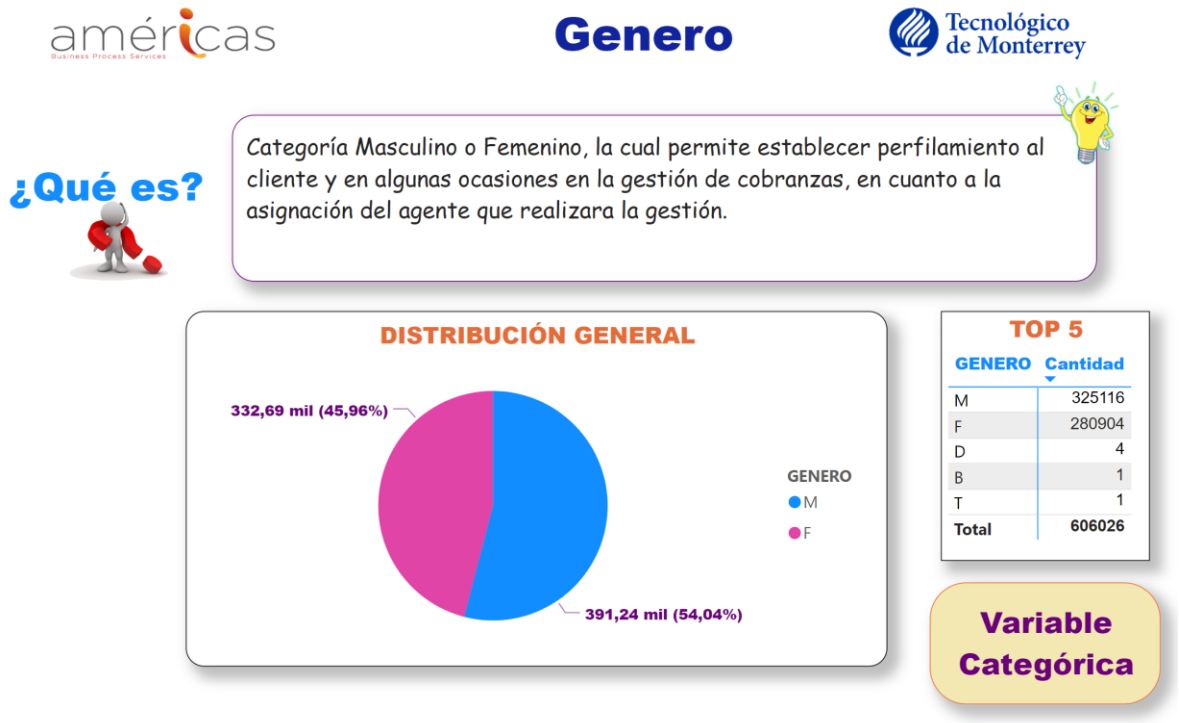
Actividad económica



Esta es una característica importante para el desarrollo de estos modelos analíticos, ya que en el proceso de cobro la actividad económica permite entender más sobre la realidad del cliente y tener un mejor perfilamiento de su comportamiento financiero. Aunque, aquí encontramos un gran reto y es que este es un dato solicitado por la entidad financiera en formatos de actualización de datos y aquí en Colombia se maneja un código de actividad económica, el cual puede llegar a ser bastante especifico, que bajo un escenario donde tengamos los datos proporcionalmente distribuidos funcionaria muy bien, pero en este caso NO es así, ya que hay una fuerte tendencia en la categoría asalariados, y esta constituye a información que puede no ser del todo precisa ya que puede que el cliente en el desconocimiento de no saber su actividad económica, diligencia en ese campo que tiende a ser general y entendido como “Alguien que tiene trabajo y recibe dinero a cambio de el”, aunque si lo tomamos a favor y evitamos un sobredimensionamiento de esta característica, al momento de realizar la ingeniería de características se propondrá agrupar en categorías mas generales y de esa forma no sobrecargar el entrenamiento del modelo, y para manejar los valores vacíos por defecto se asignaran a la categoría general de asalariados, bajo la premisa previamente mencionada, así garantizamos una apropiada distribución en las

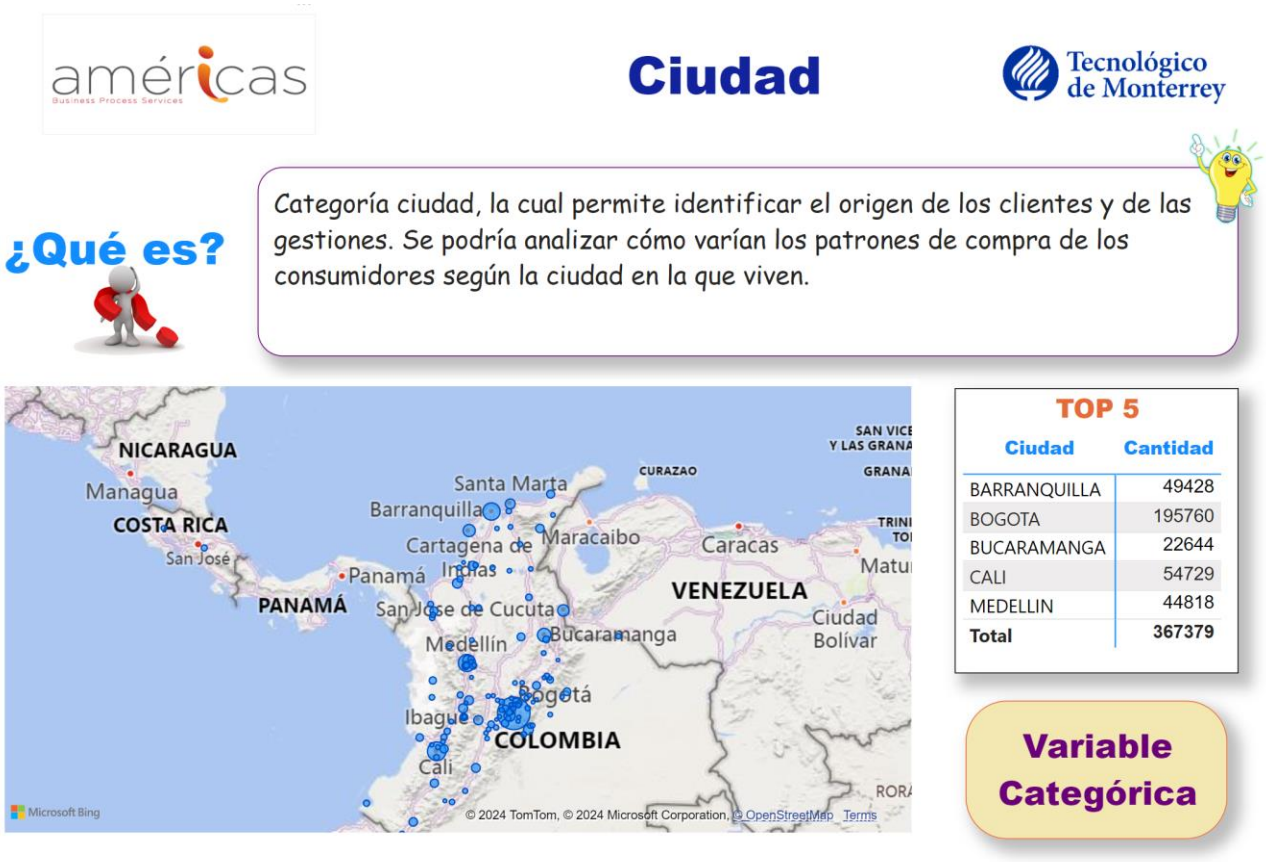
categorías y una mejor clasificación para el modelo, la cual se ira ajustando en que tan generales se dejaran las categorías y haciendo algunos experimentos en cual funciona mejor.

Genero



Esta característica, permite perfilar el tipo de base de gestión que se recibe, si bien el género solo no permite identificar tendencias concretas sobre la probabilidad de que un cliente dedica ponerse o no al día con su obligación financiera, si puede ayudar a marcar tendencias al verlo con su historial financiero y demás elementos. En ocasiones se observan otras categorías mencionadas con letras D,B,T y se refiere a que en el momento de diligenciar el formulario de actualización de datos o el cliente decide no llenar ese campo o lo deja como no especificado, que si bien no son muchos registros, estas categorías si pueden ser de gran ayuda para tratar con valores vacíos ya que en las técnicas actuales por el solo contenido semántico del nombre no podríamos asumir el género de una persona así que si se llega a dar el caso se dejaran con la letra D. Adicionalmente, podemos observar que hay una distribución similar en cuanto a los clientes siendo un 10% predominante la cantidad de hombres vs la de mujeres.

Ciudad



El análisis de datos por ciudad permite descubrir patrones y tendencias específicos de ciertas ubicaciones. Esto puede resultar útil para comprender las diferencias regionales, identificar áreas de concentración o evaluar el impacto de factores locales en las variables de interés. La variable "ciudad" puede ser una herramienta poderosa para descubrir patrones geográficos, segmentar datos, comparar ciudades, visualizar relaciones espaciales y comprender el contexto local en diversas aplicaciones de análisis de datos. Por ejemplo, en este gráfico se puede visualizar que la ciudad de Bogotá concentra el máximo de gestiones de la empresa, de hecho, concentra el mayor porcentaje de gestiones con respecto del total.

Al comparar datos de diferentes ciudades, puede obtener información sobre el rendimiento relativo, las características o las tendencias de cada ubicación. Esto puede resultar útil para realizar evaluaciones comparativas, identificar valores atípicos o comprender el impacto de factores específicos de la ciudad. Al analizar datos que involucran comportamiento humano,

preferencias o factores socioeconómicos, la variable ciudad puede proporcionar contexto sobre el entorno local y las influencias que pueden dar forma a los datos.

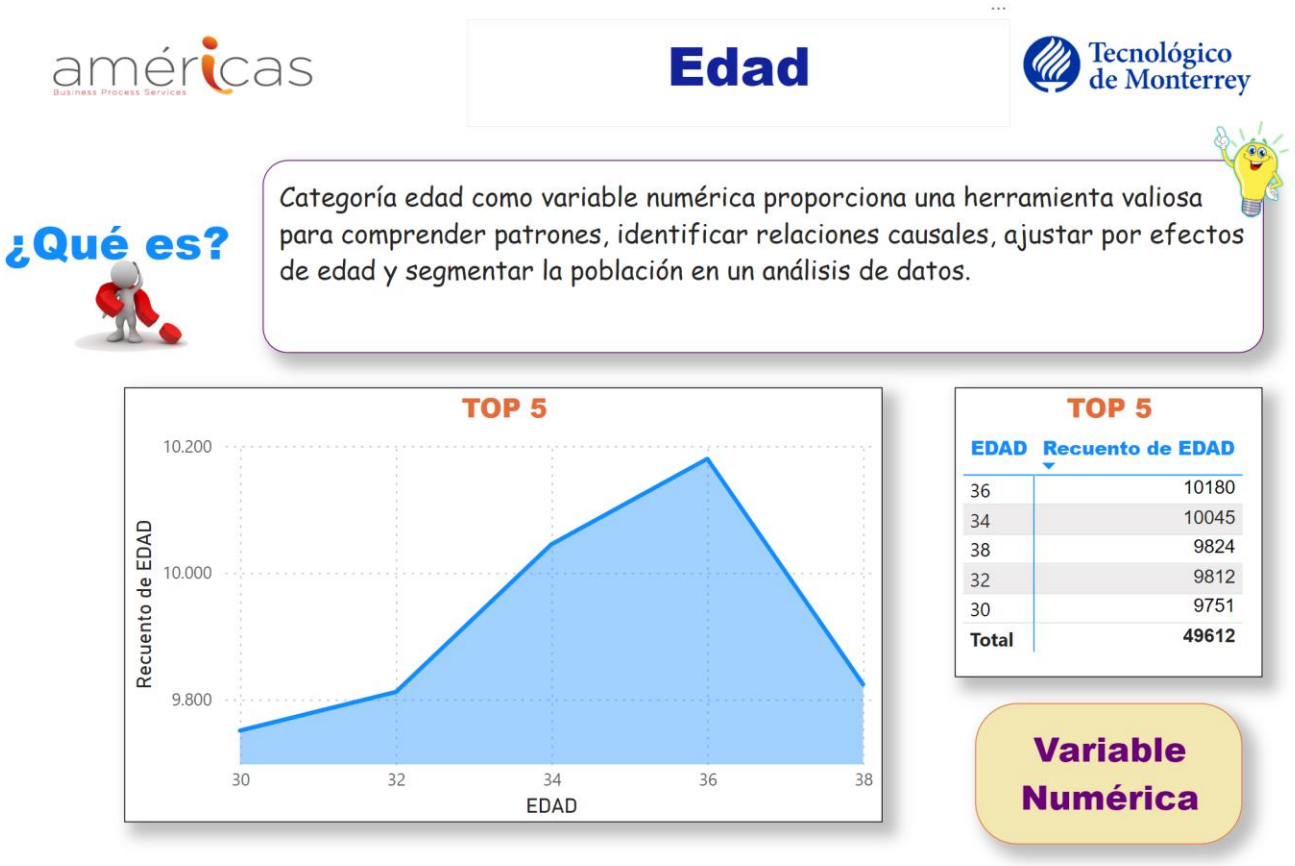
Desc-actividad

La descripción de la actividad económica como variable categórica es una herramienta valiosa para el análisis de datos en diversos ámbitos. Al comprender los patrones y tendencias asociados con esta variable, se pueden obtener conocimientos importantes que pueden conducir a mejores decisiones de inversión, políticas públicas, estrategias empresariales y comprensión del panorama económico y social.

Es importante destacar que la utilidad de la DAE como variable dependerá en gran medida del contexto específico del análisis. Se debe considerar cuidadosamente la forma en que se define y se clasifica la variable, así como las otras variables relevantes en el conjunto de datos, para garantizar la obtención de conclusiones precisas y significativas. Considerando el gráfico podemos observar las principales actividades económicas por ciudad, como el comercio al por menor la segunda actividad más grande y asalariados la tercera, mientras que Bogotá concentra más actividades económicas con respecto al resto de ciudades de Colombia.

Se puede usar para analizar la estructura y el desempeño de la economía de un país o región, identificando los sectores económicos más importantes, las tendencias de crecimiento y las áreas que requieren atención.

Edad



La variable "edad" es una herramienta poderosa para descubrir patrones relacionados con la edad, segmentar datos, predecir resultados, evaluar los efectos de la edad y comprender los eventos del ciclo de vida. Al utilizar eficazmente esta variable, puede obtener conocimientos más profundos sobre el comportamiento humano, las tendencias demográficas y los fenómenos relacionados con la edad en varios campos. En este gráfico podemos observar que las ciudades con un promedio de edad mayor son Bogotá, Cali y Barranquilla, lo que nos ayuda a diferenciar la estrategia para correlacionar los resultados con las edades de nuestros clientes. El análisis de los datos de ventas por edad puede revelar preferencias generacionales, patrones de uso de productos o la eficacia de las campañas de marketing dirigidas a grupos de edad específicos.

La edad puede ser un predictor importante de diversos resultados, como el comportamiento del consumidor, los riesgos para la salud, el nivel educativo o la participación en el mercado laboral. Al analizar patrones relacionados con la edad, se pueden desarrollar modelos predictivos para pronosticar tendencias o comprender la probabilidad de ciertos resultados.

Adicional, debemos examinar con mucho cuidado esta variable, ya que tenemos valores que no cuentan con ese registro de edad “Real” y se asigna por defecto una fecha genérica como lo es 01/01/1900, esto ocasiona que al generar algún tipo de grafico tengamos desviaciones significativas y esto genere un sesgo en los datos, por lo tanto, para la parte de ingeniería de características se revisara y se garantizaran edades que solo esten en el rango de 18 a 80 años.

Mora Inicial

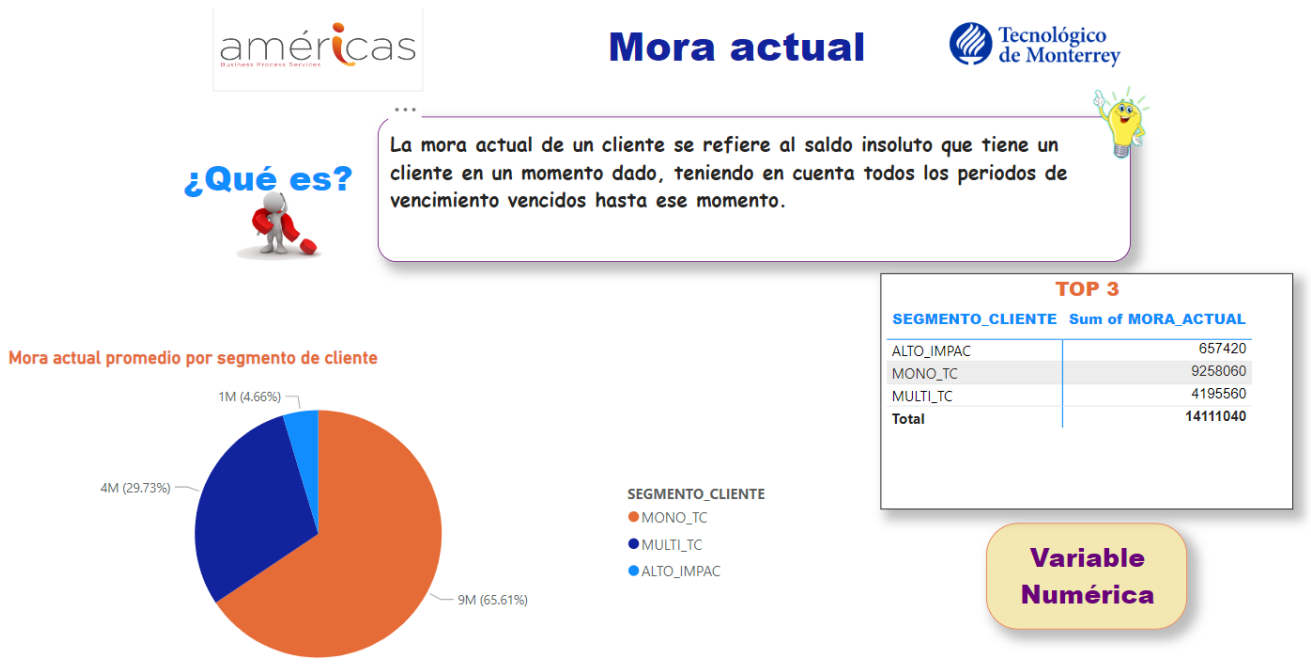
La mora inicial se define como el número de días que transcurren desde la fecha de vencimiento de un pago hasta la fecha en que se realiza el mismo. En el contexto de un contact center de Américas, la mora inicial es utilizada para evaluar la eficiencia del proceso de cobranza y la capacidad de los clientes para cumplir con sus obligaciones financieras. La mora inicial puede ser utilizada para realizar diversos análisis:

- Identificar a los clientes con mayor riesgo de morosidad: Al analizar la distribución de la mora inicial, es posible identificar a los clientes que tienen mayor probabilidad de

incumplir con sus pagos. Esta información se usa para implementar estrategias de cobranza preventiva.

- Evaluar la efectividad de las estrategias de cobranza: Al comparar la mora inicial antes y después de implementar una nueva estrategia de cobranza, es posible evaluar la efectividad de esta.
- Segmentar a los clientes: La mora inicial puede ser utilizada para segmentar a los clientes en función de su comportamiento de pago. Esta información se utiliza para desarrollar estrategias de cobro en este caso.

Mora actual



Se define como el número de días desde la fecha de vencimiento del pago más reciente hasta el análisis y se usa para evaluar la situación actual de la cartera de morosos y la efectividad de las estrategias de cobranza, es una variable numérica que puede tomar valores enteros positivos. Valores más altos indican que los clientes demoran más tiempo en pagar, se puede crear un reporte de morosidad que identifique a los clientes que requieren atención inmediata y desarrollar modelos de riesgo que permitan identificar a los clientes con mayor probabilidad de incumplir con sus pagos en el futuro. Esta variable nos ayuda analizar los siguientes puntos:

- Al comparar la mora actual en diferentes periodos de tiempo, es posible evaluar la tendencia de la morosidad en el contact center. Esta información puede identificar problemas en el proceso de cobranza o en la cartera de clientes.
- Evaluar la efectividad de las estrategias de cobro: Al comparar la mora actual antes y después de implementar una nueva estrategia de cobro, es posible evaluar la efectividad de esta.

Segmento de Cliente

Es una variable categórica que clasifica a los clientes en grupos en función de características compartidas. Estas características pueden ser demográficas, geográficas, psicográficas o de comportamiento, se utiliza para comprender mejor las necesidades y preferencias de los clientes, y para desarrollar estrategias de marketing y ventas más efectivas. podría transformarse en una variable numérica para enriquecer el análisis y mejorar el rendimiento del modelo.

Las siguientes consideraciones son importantes al utilizar la variable "segmento de cliente" categórica:

- Preserva la información cualitativa: Permite mantener la información original sobre las categorías del segmento de cliente, lo que puede ser útil para comprender mejor los patrones en los datos.
- Facilita la interpretación de los resultados: La interpretación de los coeficientes del modelo de machine learning es más directa y transparente cuando se utiliza una variable categórica.
- Puede aumentar la dimensionalidad del modelo: La codificación one-hot encoding puede aumentar significativamente la dimensionalidad del modelo, lo que puede generar problemas de sobreajuste y dificultar la interpretación de los resultados.
- No permite realizar algunas operaciones matemáticas: Al ser una variable categórica, no es posible realizar ciertas operaciones matemáticas, como calcular promedios o desviaciones estándar.

Tipo de Cliente

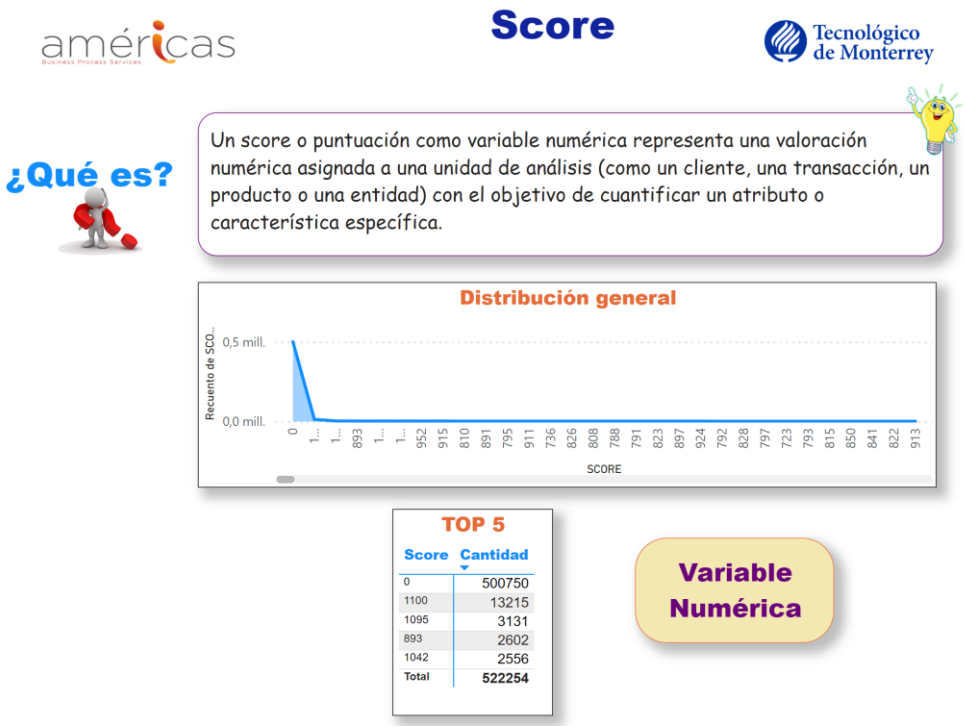
Es una variable categórica para la clasificación de los clientes según características como su perfil de consumo, hábitos de pago, historial de cobro o cualquier otra característica relevante para el análisis, y para identificar patrones que ayuden a determinar el horario más efectivo de cobranza. Esto significa que los valores de la variable son categorías cualitativas que no tienen un orden natural. En nuestro caso, estamos evaluando enviar una campaña de cobro por correo electrónico a los clientes nuevos, y se puede llamar por

teléfono a los clientes morosos, por otro lado, también es útil para identificar a los clientes con mayor riesgo de morosidad antes de retrasar sus pagos.

El cliente puede usarse para determinar el horario más efectivo de cobranza para cada grupo de clientes. Por ejemplo, se puede observar que los clientes nuevos son más receptivos a las llamadas de cobranza por la mañana, mientras que los clientes frecuentes son más receptivos por la tarde

Es importante tener en cuenta que la segmentación de clientes por tipo de cliente es solo uno de los factores que se deben considerar al determinar el horario más efectivo de cobranza. Otros factores importantes incluyen la zona horaria del cliente, sus hábitos de trabajo y sus preferencias de comunicación

Score



Es una puntuación numérica asignada a cada cliente en función de su riesgo de morosidad o su probabilidad de pago exitoso, Valores más altos de score indican un menor riesgo de morosidad y una mayor probabilidad de pago exitoso. Por el contrario, valores más bajos de score indican un mayor riesgo de morosidad y una menor probabilidad de pago exitoso. Nos puede ayudar para mejorar las analizar las siguientes consideraciones:

- Priorizar las acciones de cobranza: Al ordenar a los clientes por su score, se puede priorizar las acciones de cobranza, comenzando por los clientes con mayor riesgo de morosidad.
- Evaluar la efectividad de las estrategias de cobro: Al comparar la morosidad antes y después de implementar una nueva estrategia de cobro para un grupo de clientes con un rango de score específico, es posible evaluar la efectividad de la misma.

Se puede crear un grupo de clientes con score alto (bajo riesgo de morosidad), un grupo de clientes con score medio y un grupo de clientes con score bajo (alto riesgo de morosidad).

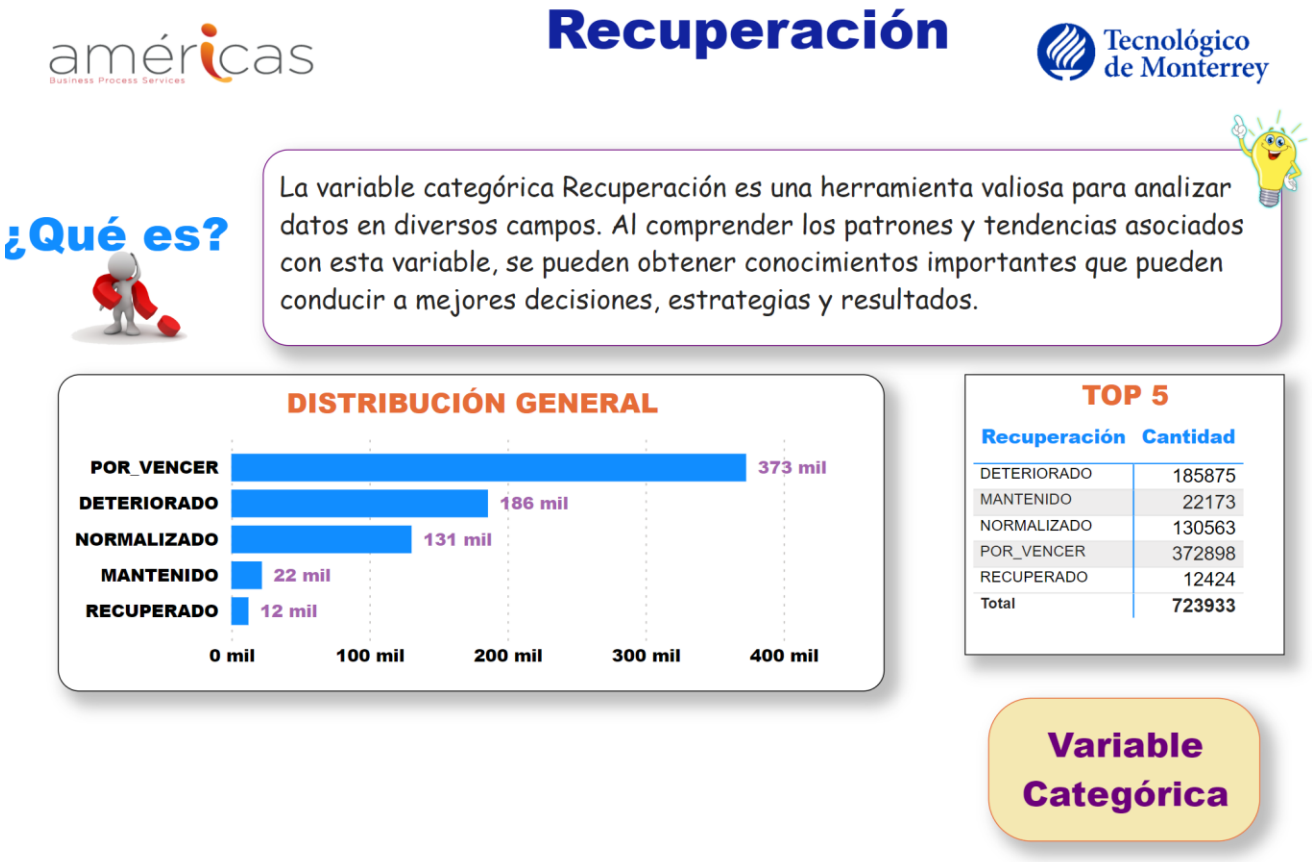
Consideramos que nos puede ayudar a determinar el horario más efectivo de cobranza para cada grupo de riesgo. Por ejemplo, se podría observar que los clientes con score alto son más receptivos a las llamadas de cobranza por la mañana, mientras que los clientes con score bajo son más receptivos por la tarde

Es importante que tengamos en cuenta las siguientes consideraciones al utilizar esta variable:

- El método utilizado para calcular el score debe seleccionarse cuidadosamente para que sea preciso y confiable posible.
- Es importante monitorear el desempeño del score a lo largo del tiempo y realizar ajustes según sea necesario.
- Se deben tomar medidas para evitar sesgos en el cálculo del score, como sesgos raciales o de género.

En adición, como podemos observar en la gráfica, la distribución del Score es demasiado amplia, no tenemos grupos grandes con una misma puntuación, lo cual implica que para la ingeniería de características debamos pensar en asignar grupos y así evitar un sobredimensionamiento de esta variable, adicionalmente, se tienen algunos valores negativos que pueden ser un error en la base y debamos revisar y adaptar.

Recuperación



La variable recuperación se refiere a si un cliente ha realizado un pago exitoso o no después de un intento de cobranza, Los valores de la variable proporcionan información sobre el éxito o fracaso de los intentos de cobranza y nos ayudan a Identificar el horario con mayor tasa de recuperación de pagos, al analizar la tasa de recuperación de pagos por hora, se puede identificar el horario en el que los clientes son más propensos a realizar un pago exitoso así como evaluar la efectividad de las estrategias de cobro, al comparar la tasa de recuperación de pagos antes y después de implementar una nueva estrategia de cobro, es posible evaluar la efectividad de la misma.

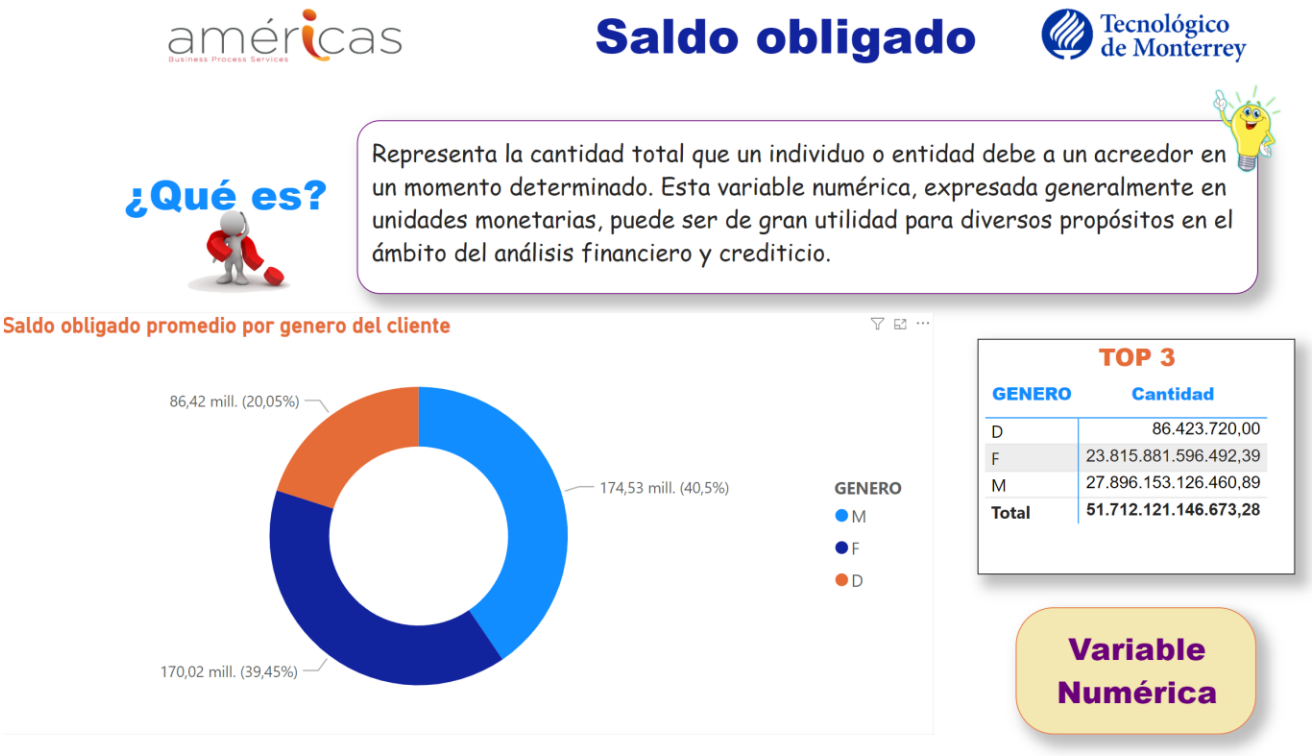
El objetivo de usar esta variable se basa en programar las llamadas de cobranza en el horario con mayor tasa de recuperación de pagos desarrollando un sistema de asignación de llamadas de cobranza, que podría asignarlas a los agentes según el horario del día y del tipo de cliente, para maximizar la probabilidad de recuperación, al comparar la tasa de

recuperación de cada agente, identificar a quienes tienen un mejor desempeño y brindar capacitación adicional a quienes lo necesitan.

Es importante que tengamos en cuenta las siguientes consideraciones al utilizar esta variable:

- Es importante definir claramente la definición de "recuperación" en el contexto del análisis.
- Se deben considerar diferentes métodos para codificar las categorías de la variable "recuperación" (por ejemplo, binaria, ordinal).
- Se deben tomar medidas para evitar sesgos en la recolección y análisis de los datos de recuperación.

Saldo Obligado

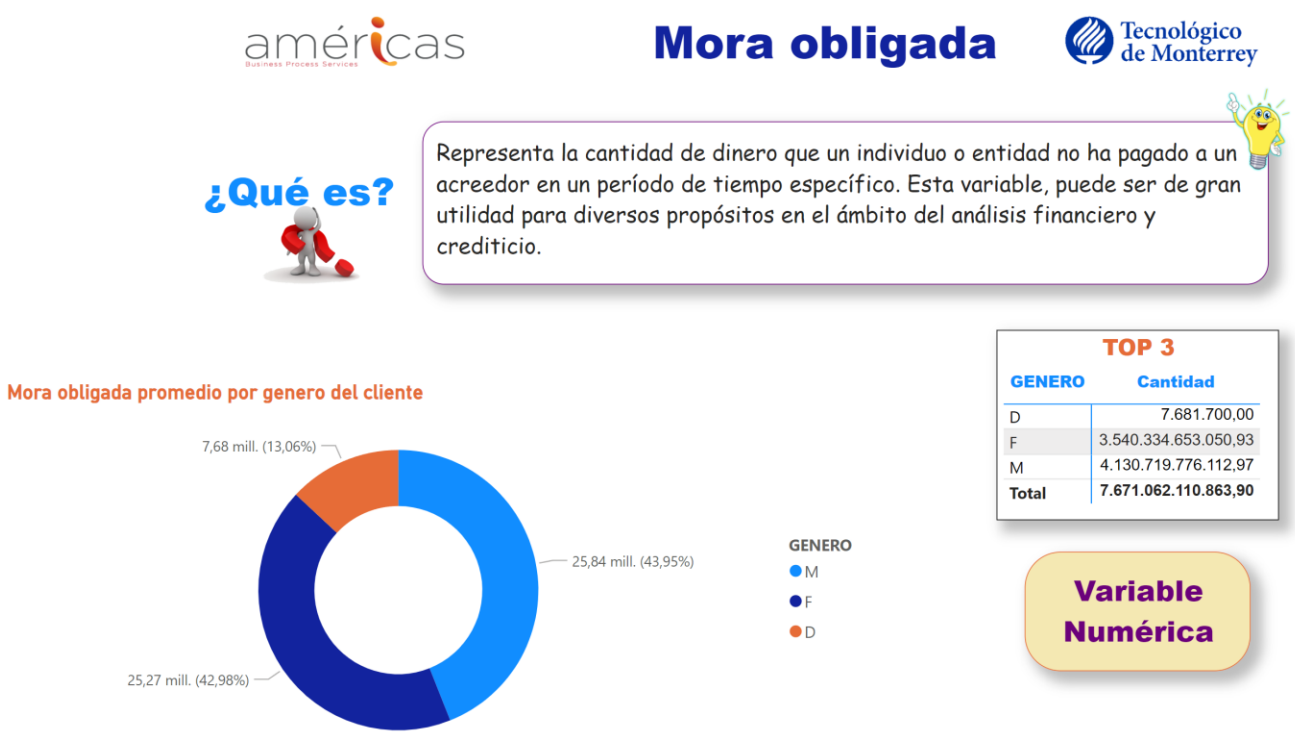


El saldo obligado puede ser una herramienta valiosa para el análisis de negocios en diversas áreas, ya que proporciona información crucial sobre la situación financiera de una empresa

o entidad. El saldo obligado puede ser un indicador de la capacidad de una empresa para cumplir con sus obligaciones financieras a corto y largo plazo. Un alto saldo obligado en relación con los ingresos o el patrimonio neto puede sugerir un mayor riesgo de insolvencia.

El saldo obligado es una variable financiera crucial que proporciona información valiosa para diversos aspectos del análisis de negocios. Al comprender y analizar el saldo obligado, las empresas y los analistas pueden tomar decisiones informadas que mejoren la salud financiera, la rentabilidad, el valor y las perspectivas de crecimiento a largo plazo. En el gráfico podemos ver el saldo obligado promedio por genero del cliente, lo anterior para generar una mejor estrategia financiera que nos sirva para recuperar la inversión lo más rápido posible.

Mora Obligada



Se refiere al tiempo que un cliente ha estado en mora con el pago de una deuda. Esta variable puede ser útil para evaluar el riesgo de morosidad y para identificar patrones que puedan ayudar a determinar el horario más efectivo de cobranza.

La mora obligada puede ser un indicador importante de la salud financiera de una empresa o entidad, así como de su comportamiento de pago. Forma parte del historial crediticio de

una empresa, lo que lo convierte en un factor crucial para evaluar su riesgo crediticio. Un historial de moras frecuentes puede indicar dificultades financieras y un mayor riesgo de incumplimiento en el futuro. En el gráfico se puede visualizar la mora promedio por genero de nuestros clientes.

Marca Tipo Cartera

Esta categorización puede ser útil para diversos propósitos en el análisis de negocios, incluyendo:

- Segmentación de clientes: La marca de tipo de cartera permite segmentar el mercado de consumidores de carteras en grupos con necesidades y preferencias específicas. Esto facilita el desarrollo de estrategias de marketing y ventas dirigidas a cada segmento.
- Análisis de tendencias: Al rastrear las tendencias en las marcas de tipo de cartera más populares, las empresas pueden identificar oportunidades de mercado emergentes y adaptar sus productos y servicios en consecuencia.

- Diferenciación de productos: Al comprender las marcas de tipo de cartera existentes, las empresas pueden diseñar productos que se diferencien de la competencia y atraigan a segmentos específicos de consumidores.

Referencias:

- Américas BPS, Qué hacemos, enlace: https://www.americasbps.com.co/que-hacemos/
- Acurio Armas, J. A., Álvarez Gómez, L. K., Manosalvas Gómez, L. R., & Amores Burbano, J. E. (2020). Modelo de gestión del talento humano para la empresa Contigo S.A. del Cantón Valencia,Ecuador. Revista Universidad y Sociedad, 12(4), 93–100.
- Adnan, A. Z., Ahman, E., Disman., Yuniarsih, T., Fattah, N., Suwatno., & Hadi Senen, S. (2022).Model Of Employee Performance Development Based On Talent Management At Pt Pertamina Ru-Vi Balongan Indramayu West Java. Journal of Positive School Psychology, 6(6), 8960–8970.
- Behrens, J. T. & Yu, C. H. (2003). Exploratory data analysis. In J. A. Schinka & W. F. Velicer, (Eds.), Handbook of psychology Volume 2: Research methods in Psychology (pp. 33-64). New Jersey: John Wiley & Sons, Inc.
- Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. Psychological Methods, 2, 131-160.
- Martinez, W. L. (2005). Exploratory data analysis with MATLAB. London: Chapman & Hall/CRC.
- T. V. Rao. (2016). Performance Management: Toward Organizational Excellence: Vol. 2nd edition. Sage Publications Pvt. Ltd. http://0-search-ebscohost-com.biblioteca-ils.tec.mx/login.aspx%3fdirect%3dtrue%26db%3dnlebk%26AN%3d1234051%26lang%3des%26site%3deds-live%26scope%3dsite%26ebv%3DEB%26ppid%3Dpp_157
- Ralph W. Adler. (2018). Strategic Performance Management : Accounting for Organizational Control. Routledge. http://0-search-ebscohost-com.biblioteca-ils.tec.mx/login.aspx%3fdirect%3dtrue%26db%3dnlebk%26AN%3d1714689%26site%3dehost-live%26ebv%3DEB%26ppid%3Dpp_i
- Son, J., Park, O., Bae, J., & Ok, C. (2020). Double-edged effect of talent management on organizational performance: the moderating role of HRM investments. International Journal of Human Resource Management, 31(17), 2188–2216. https://doi.org/10.1080/09585192.2018.1 443955
- Tukey, J. W (2009). Exploratory Data Analysis as part of a larger whole. In L. V. Jones (Ed.), The collected works of John W. Tukey: Vol. IV. Philosophy and principles of data analysis: 19651986 (pp. 793803). Pacific Grove, CA: Wadsworth. (Original work published 1973).
- Wolor, C. W., Khairunnisa, H., & Purwana, D. (2020). Implementation talent management to improve organization's performance in Indonesia to fight industrial revolution 4.0. International Journal of Scientific and Technology Research, 9(1), 1243–1247.
- Zulqurnain, A., Madeeha, B., & Aqsa, M. (2019). Managing Organizational Effectiveness through Talent Management and Career Development: The Mediating Role of Employee Engagement. Journal of Management Sciences, 6(1), 62–78. https://doi.org/10.20547/jms.2014.1906105