UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE CIENCIAS ECONÓMICO ADMINISTRATIVAS

COORDINACIÓN DE POSGRADO

**MAESTRÍA EN CIENCIA DE LOS DATOS**



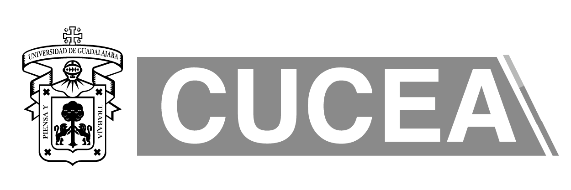
Proyectos 1

Modelo de selección de productos clave en canal tradicional

Presenta

Miguel Ángel López Rojas

Octubre del 2023



**Resumen**

El presente proyecto, está enfocado en a la búsqueda del mercado ideal para la industria abarrotera.

Al ser un mercado saturado y con más de 88,000 SKU’s en el mercado nacional, la presente tesis busca tanto el Estado de la Republica menos saturado, hasta la categoría más atractiva a enfocarse.

Para esta tesis se emplea una serie de técnicas y herramientas, desde Geopandas, hasta modelos Montecarlo para predecir los futuros en la industria.

Las fuentes de información para los mismos, son desde mapas de INEGI, datos de la misma entidad, ISCAM en archivos de Excel y las metodologías fueron técnicas utilizadas en las clases de la maestría en Ciencias de datos.

**Palabras clave:** Abarrotes, Modelo de selección, Montecarlo, Regresión Lineal, Productos clave, Mercado nacional

**Def****inición**

El canal tradicional, refiera a la comercialización alimentos, bebidas y artículos de higiene y limpieza, también llamado canal abarrotero.

Son productos básicos para las rutinas diarias, que suelen referirse a la cadena desde productor (proveedor) hasta la tienda de abarrotes, por lo que gestiona una alta cantidad de productos, proveedores y vendedores, que suelen tener bajos márgenes de utilidad, debido a la alta competencia y alta necesidad de dichos productos en la vida diaria de las personas.

**Planteamiento del problema**

Debido a la naturaleza de la saturación en el mercado, existe una baja rentabilidad, amplia cadena de suministro (Generalmente siendo: Fabricante, revendedor mayoreo, revendedor minorista y abarrotera a cliente final), limitados espacios para sus inventarios y por último saturación del mercado y empresas abarroteras mayoristas, se buscará a través de la presente tesis el encontrar los productos y Estado de más atractivos del canal, para concentrar la operación en artículos y mercados clave.

Un amplio portafolio, significa también baja rotación de inventarios, lo que implica costos y baja en flujo de efectivo. La especialización de productos es una ventaja competitiva al concentrar los esfuerzos en pocos objetivos.

**Justificación**

Las casas comerciales de consumo masivo

Son las grandes fábricas que se encargan de la producción o elaboración de los productos de consumo masivo. Los productos de consumo masivo o de alta demanda son aquellos productos requeridos por todos los estratos de la sociedad para la vida diaria, que se distribuyen en categorías como: lácteos, bebidas, tocador y aseo personal, aseo, confitería y snacks, panadería, entre otras.

La labor de estas compañías consiste en realizar todo el proceso que conlleva sacar un producto al mercado, desde la idea inicial, como el análisis de mercado, las pruebas, la producción, el empaquetamiento, etc.

Lo que motiva gran parte de la competencia entre estas empresas es el aumento en clientes, el aumento en territorio (cubrir cada vez más tiendas o puntos de venta).

Las tiendas de barrio, changarros o bodegas

Son los [pequeños establecimientos](https://business-intelligence.grupobit.net/blog/radiografia-de-una-tienda-de-barrio), que normalmente son atendidos por una o dos personas y a menudo son negocios familiares. Venden productos de consumo masivo o de la canasta familiares a quienes, normalmente, viven cerca del establecimiento. Dentro de la categoría de tiendas de barrio se pueden considerar también a las farmacias, panaderías y misceláneas.

En un principio, estas tiendas solían ser el lugar de compra para productos como bebidas, botanas y golosinas, sin embargo, se ha transformado significativamente incluyendo ahora productos de uso doméstico, higiene, etc.

Las tiendas de barrio son un eje fundamental en la economía de los países latinoamericanos, pues son una fuente importante de generación de empleo y porque llegan a los lugares más apartados de los países.

[Conozca algunos consejos para que las casas comerciales y los distribuidores mejoren la relación con sus tenderos](https://business-intelligence.grupobit.net/blog/4-consejos-para-mejorar-la-relacion-con-sus-tenderos)

Los distribuidores

Los distribuidores son el eslabón intermedio para que en el canal tradicional los productos lleguen de la fábrica que los produce a cada tienda y por ende a los hogares.

Las casas comerciales trabajan en alianza con los distribuidores pues son ellos los que tienen los medios (camiones, carros, etc.), la experiencia en distribución y la logística para llevar los productos de consumo masivo a todos los rincones del territorio, incluso los más apartados.

El distribuidor apuesta por determinados productos o compañías de consumo masivo en función de su experiencia en el canal tradicional, sabiendo qué es lo que más podría venderse en cada área del país o de la región.

Una vez la empresa factura y recibe el pago del distribuidor, este es quien se hace responsable de los productos y se encarga de venderlos por cuenta propia, estableciendo el precio, frecuencia y lugares donde llegan.

**Antecedentes/Marco teórico**

El canal tradicional es el formato de venta más antiguo en Latinoamérica, remontando sus raíces en los tiempos de la colonia, cuando se da la aparición de pequeñas tiendas a donde las personas acudían a comprar sus alimentos periódicamente. Desde entonces, estas se han convertido no solo en canal de compra favorito en Latinoamérica sino en un lugar simbólico para las comunidades, convirtiéndose también en un espacio de reunión y socialización.

Más adelante, para el siglo XX empezaron a aparecer nuevos formatos de tienda como los supermercados o los minimercados, como respuesta al aumento en las poblaciones, por tanto, aumento en la demanda y al aumento en la oferta de productos y marcas de consumo masivo. Estos nuevos formatos se encargaron de industrializar y sistematizar el proceso de venta de productos de consumo masivo, utilizando grandes instalaciones, pasillos y góndolas, después agregando cajas registradoras y modernizando sus instalaciones. Este tipo de tiendas las conocemos hoy en día como el canal moderno.

No obstante, a pesar de más de 5 siglos de antigüedad, de cientos de miles de tiendas a lo largo de todo el continente latinoamericano y la aparición de nuevos y modernos formatos de tiendas como los “Hard discounters” o supermercados por membresía, sigue siendo el canal donde se realizan aproximadamente 60% de las compras de consumo masivo del continente.

**Hipótesis**

A través de conocer las tendencias del mercado, crecimientos, participaciones y valores de los mismo, podremos conocer los mercados más interesantes del canal y selección de productos más rentables.

**Objetivos**

Facilitar herramientas para la elección y comercialización de productos a través de un modelo predictivo que permita seleccionar los artículos más interesantes para el crecimiento, selección de los mismos, así como su mercado.

**Metodología**

Para este proyecto, lo primero fue la descarga de las empresas mayoristas en México en INEGI, así como la descarga de Mapas que sirvan como delimitadores de los puntos geográficos de estas empresas. El fin de esto es encontrar la menor saturación posible mediante Geopandas y el uso de DBSCAN como método espectral para conocer los espacios entre los puntos.

Una vez identificado el Estado de la Republica más atractivo.

Sobre un formato de ISCAM (El cuál mide los crecimientos del mercado abarrotero) creamos un modelo Montecarlo para predecir en base a los crecimientos de cada categoría con una distribución de Peart, pues las colas pueden ser muy largas, tomamos como valor más probable el crecimiento actual de la categoría, como mínimo el porcentaje de la categoría que menos creció y máximo, el porcentaje de la categoría que más creció.

Para simular el Estado más atractivo, que resultó en el DBSCAN ser Durango, como ISCAM no viene dividido por Estados y es un general de mercado, dividimos el total de empresas de INEGI (168040) entre el número de empresas en Durango (7077), lo cuál nos dio un 4.21%, mismo que le asignamos al peso del mercado para realizar también un modelo Montecarlo para simular el crecimiento del mercado en Durango.

**Recursos**

A través de las bases de datos de ISCAM, podremos conocer las tendencias que marcarán dicho estudio, al inicio de la presente tesis, nos encontramos haciendo un acercamiento con dicha empresa que se encarga de recopilar los datos de las empresas proveedoras y distribuidas del canal.

ISCAM es una empresa SA de CV, por lo que dicha base de datos es de paga, mientras la negociación se realiza, se realizará el modelaje con una base de datos sintética.

También utilizamos la base de datos del INEGI, tanto para la división territorial, como la base de datos de las empresas abarroteras en México

**De lo general a lo particular:**

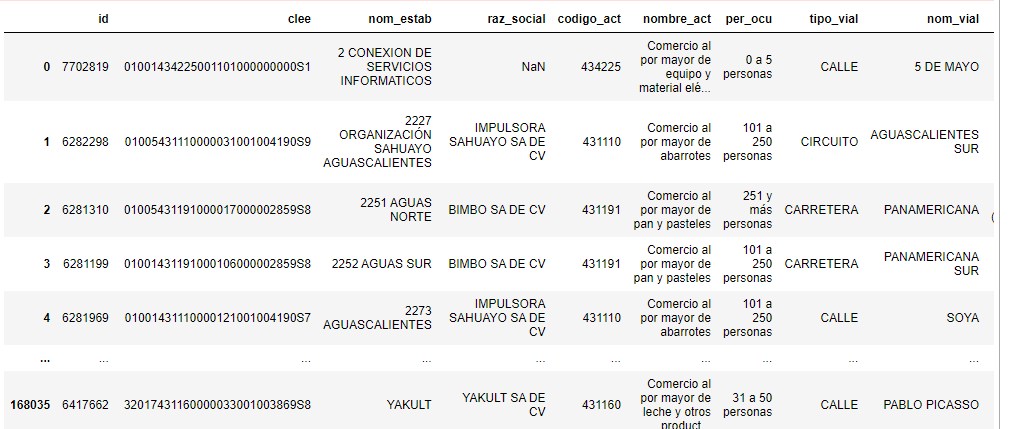
El primer paso es la búsqueda del estado de la Republica con menor saturación en base a cantidad de empresas por kilómetro cuadrado.

Descargamos de la base del INEGI que contiene la información de cuantas empresas mayoristas existen en México, así como el archivo .shp para delimitar los espacios de cada Estado.

El archivo que contiene las empresas se descarga y se transforma en csv para una lectura más sencilla.

A Continuación, se transforma a DataFrame de Pandas en una variable que se llamará df:

df = pd.read\_csv('denue\_inegi\_43\_.csv', encoding='ISO-8859-1')



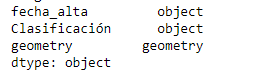
**Transformándolo a data frame de Geopandas:**

Ya que la información no contiene nombre de los estados, no podemos catalogarlas directamente por Estado, por lo que para encontrar a que estado pertenecen y luego ver cuál es el estado con menos saturación en México, tendremos que trabajar con las columnas ‘latitud’ y ‘longitud’, que incluyen el data frame, sin embargo, hay que transformarlo a puntos en el mapa, por tanto utilicé la librería ‘Geopandas’.

gdf\_mayoristas = gpd.GeoDataFrame(df, geometry=gpd.points\_from\_xy(df.longitud, df.latitud))

Este código convirtió la información en puntos geográficos, por lo que ahora podemos trabajar con las coordenadas que corresponden a cada estado

gdf\_mayoristas.dtypes #Asegurarnos que incluye la figura geométrica



Ahora cargamos la información que delimitará el mapa, en base a la base de datos geoestadísticos, la cual llamaremos “fig00ent”

fig00ent = gpd.read\_file(r'C:\Users\User\Maestria\Tesis\889463770541\_s\mg2022\_integrado\00ent.shp')



Al cargarla la base de datos de empresas mayoristas, así como de la forma de los mapas, vienen en diferentes escalas por lo que las ajustamos a una misma.

Empresas mayoristas:

gdf\_mayoristas.crs = CRS.from\_epsg(4326)

Mapa:

fig00ent = fig00ent.to\_crs(gdf\_mayoristas.crs)

Y los unimos en un mismo mapa, primero uniendo ambas:

gdf\_merged = gpd.sjoin(gdf\_mayoristas, fig00ent, op='within')

Después le agregamos columnas para la geometría con latitud y longitud:

gdf\_merged['geometry'] = [Point(xy) for xy in zip(gdf\_merged['longitud'], gdf\_merged['latitud'])]

gdf\_merged = gpd.GeoDataFrame(gdf\_merged, geometry='geometry', crs=gdf\_merged.crs)

El siguiente punto, ahora que se encuentran unidos los datos de los mayoristas con el mapa es identificar cuál es el Estado más atractivo, es decir, el que tenga los vectores de empresas mayoristas con mayores separaciones entre sí.

Para poder identificar esto, utilizaremos una técnica de clusterización por densidad, por lo que nos decidimos por realizarlo a través del DBSCAN.

**¿Qué es el DBSCAN?**

El DBSCAN es una técnica o algoritmo de agrupamiento espacial de datos. Por espacial nos referimos a distancias entre los puntos o vectores.

Este algoritmo requiere de dos parámetros, el primero es llamado “eps” y se refiere a la cantidad mínima de puntos requeridos en una región. La segunda es e-vecindad que se refiere a si un punto tiene vecinos, se consideran parte del cluster, en caso de que no tenga vecinos, se considera como “ruido”.

**Aplicamos el DBSCAN para el mapa**

El primer reto que nos encontramos al aplicar el DBSCAN sobre el mapa completo es que se convierte una matriz cuadrada de 168040 x (168040 – 1), por lo que el espacio para la realización de esta técnica era ineficiente, ya que se tiene que calcular la distancia de cada uno de los vectores entre el resto de los vectores, por ejemplo, a distancia con b, a distancia con c, b distancia con a, b distancia con c y así sucesivamente…

Debido a la eficiencia computacional optamos por graficar el total de empresas para ver que estados se pueden identificar a simplemente con mayores distancias entre sí de los vectores (empresas mayoristas).

Para graficar este mapa utilizamos la librería *Folium*:

import folium

# Crear un mapa centrado en los Estados Unidos

m = folium.Map(location=[37, -95], zoom\_start=4)

# Crear un grupo para los puntos

group = folium.FeatureGroup(name='Puntos')

# Iterar sobre los puntos y añadirlos al grupo

for lat, lon in coords:

folium.CircleMarker(location=[lat, lon], radius=2, fill=True, fill\_opacity=0.7, color='blue').add\_to(group)

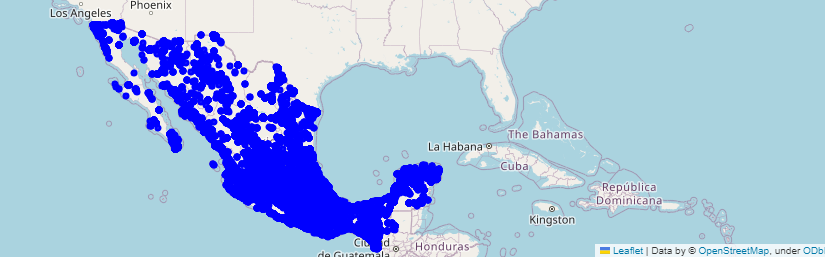
# Añadir el grupo al mapa

group.add\_to(m)

# Añadir un control para la capa de puntos

folium.LayerControl().add\_to(m)

m



Después de hacer zoom, gracias a una de las cualidades de ‘folium’, podemos apreciar que los tres estados con más espacios entre sí corresponden a Coahuila, Durango y Baja California Sur.

Ahora para hacer el proceso computacional más eficiente el proceso, generaremos el DBSCAN únicamente para estos tres Estados, por separado.

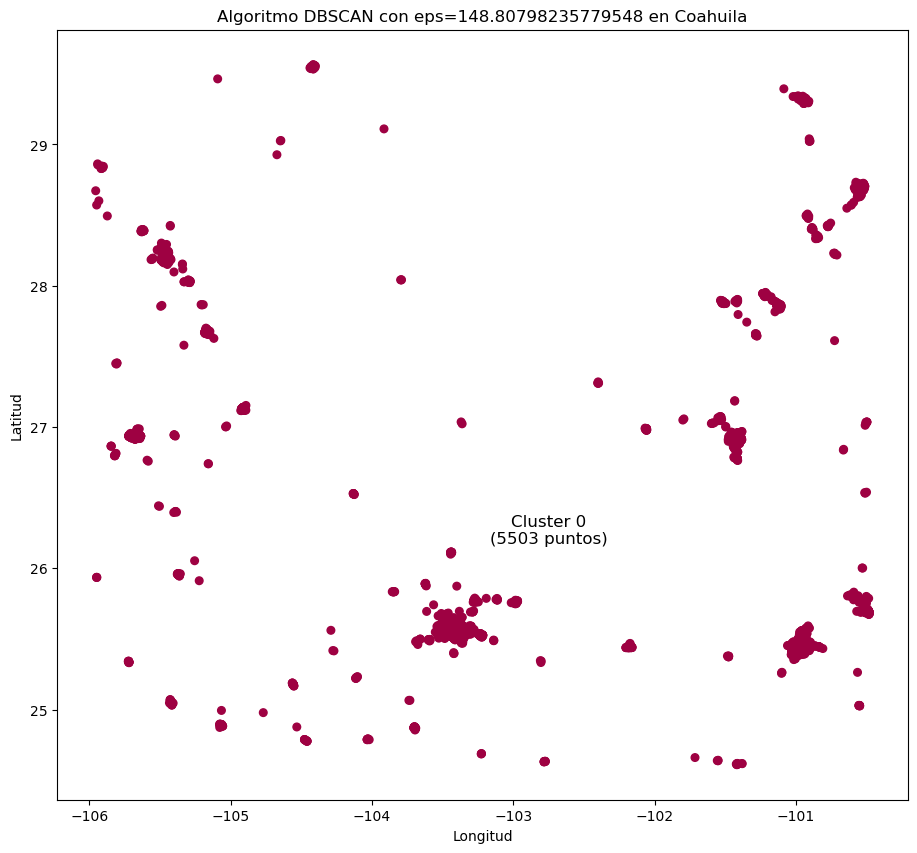
Recordamos que, para asignar un vector a cada estado, hay que especificar que delimitar las coordenadas de cada estado.

**Coahuila:**

A continuación, presento como se delimita el Estado:

coahuila\_data = gdf\_merged[(gdf\_merged['geometry'].x > -105.962142) & (gdf\_merged['geometry'].x < -100.481785) & (gdf\_merged['geometry'].y > 24.544246) & (gdf\_merged['geometry'].y < 29.583685)]

clustering\_coahuila = DBSCAN(eps=eps\_coahuila, min\_samples=5, algorithm='ball\_tree', metric='euclidean').fit(np.radians(coords\_coahuila))

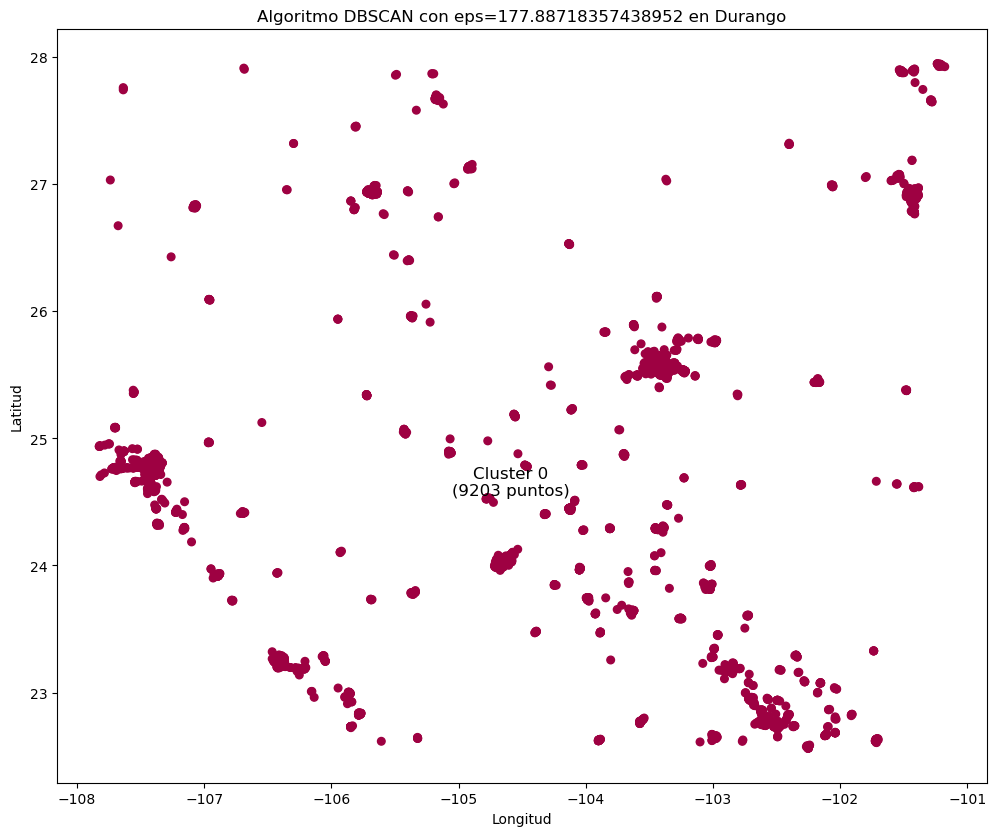


**Durango:**

El Código para delimitar las coordenadas:

durango\_data = gdf\_merged[(gdf\_merged['geometry'].x > -107.872337) & (gdf\_merged['geometry'].x < -101.170224) & (gdf\_merged['geometry'].y > 22.553077) & (gdf\_merged['geometry'].y < 27.944788)]

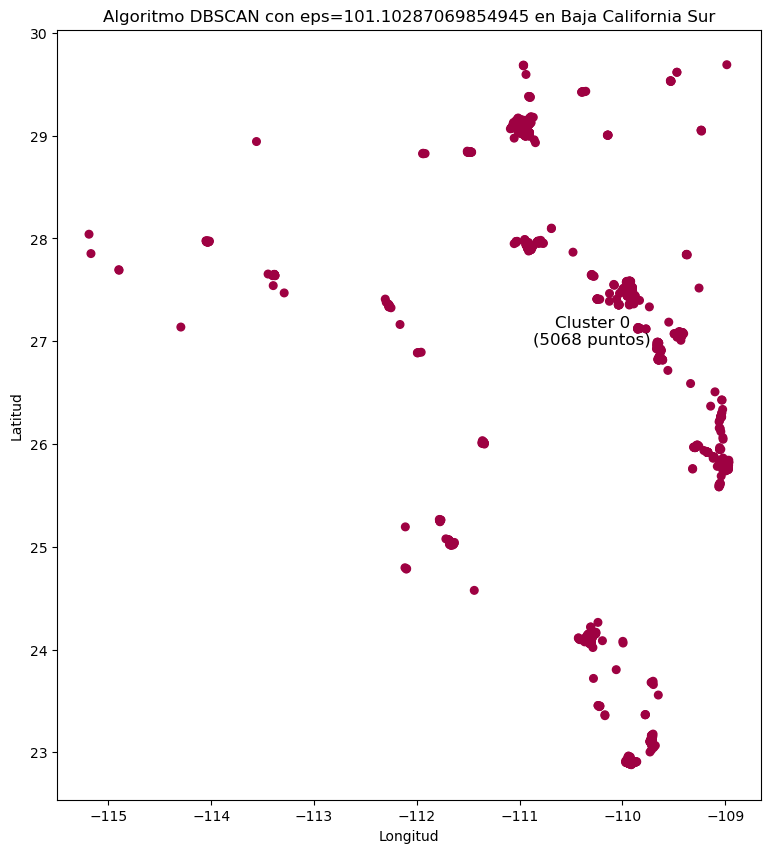
El DBSCAN de Durango:

****

**Baja California Sur:**

El código para delimitar el espacio:

bcs\_data = gdf\_merged[(gdf\_merged['geometry'].x > -115.649259) & (gdf\_merged['geometry'].x < -108.964135) & (gdf\_merged['geometry'].y > 22.318292) & (gdf\_merged['geometry'].y < 29.784869)]

****

A simple vista pareciera que la opción más atractiva por cantidad de empresas es Baja California, sin embargo, vamos a utilizar como indicador la distancia promedio entre los puntos, que nos ayudó a calcular el DBSCAN entre el número de empresas totales por estado, que calculamos en base a las delimitaciones las coordenadas usando la función len().

La ecuación es (número de empresas/ kilómetros cuadras promedio entre puntos), por tanto, entre mayor sea el resultado, más atractivo es. Los resultados fueron los siguientes:

Coahuila: 26.14872891423141

Durango: 36.58226338593632

Baja California Sur: 35.445516855504266

Lo que nos indica que el estado más atractivo a concentrarnos es Durango.

Una vez identificado que Durango es el estado más atractivo por saturación, trabajamos en un archivo con datos sintéticos de ISCAM.

**Modelo Montecarlo**

El modelo Montecarlo se crea a través de una técnica matemática que predice los posibles resultados en un evento incierto.

Como lo comentamos anteriormente se realizará una predicción del mercado, en base a un reporte sintético de ISCAM, a través de una extensión de Excel llamada RISK .

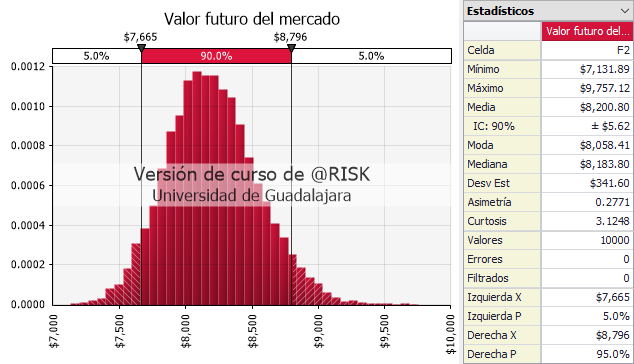
Por medio de Risk podemos predecir y calcular escenarios aleatorios en base a la forma de la distribución de los datos.

Para este ejercicio, le dimos la distribución a los escenarios con una distribución de peart, ya que la forma de peart considera colas muy amplias.

Para el valor más probable se tomaron los crecimientos reales de porcentaje, es decir, que la situación y el crecimiento se mantengan.

Como el valor mínimo se eligió el valor actual del mercado de cada categoría más el menor crecimiento en el mercado, que en este caso fue de -14%.

Por último, para el valor máximo, se tomó la categoría que más creció como referencia y se le agregó ese porcentaje, al valor actual de mercado, en este caso fue del 172%.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Categorías** | **% Crec (+/-) MPA** | **% Part. Categoría** | **Valor actual del mercado** | **Valor futuro de la categoría** |
| Galletas | 8% | 10.20% | $631.20 | $833.39 |
| Cafe | 12% | 4.90% | $304.00 | $408.47 |
| Consomes | 6% | 1.70% | $107.80 | $140.54 |
| Sueros Orales | 46% | 2.00% | $122.10 | $191.57 |
| Mayonesas | 4% | 3.10% | $192.70 | $249.29 |

A la celda del resultado de esa simulación, la convierte en color azul, una vez simulando 10,000 escenarios.

En base a estas simulaciones, tenemos la oportunidad de conocer la probabilidad de que las cosas sucedan, entonces buscamos cual es el crecimiento más probable, por medio del promedio del crecimiento de todas las categorías y le sacamos su media para este crecimiento, en este caso el promedio real de crecimiento es del 12%, sin embargo, ya con la proyección de los 10,000 escenarios en cada categoría, el crecimiento más probable es del 24.5%.

Para conocer la parte proporcional de Durango, realizamos el mismo calculo, en base a la proporción del estado en base a su número de empresas.

Por último, simulamos para el estado de Durango simulamos por este método asumiendo la misma distribución de los datos Peart y con los mismos porcentuales del mercado nacional, de mínimos y máximos.

El crecimiento de Durango es exponencial con esta forma, ya que los valores de las colas afectan de manera positiva el crecimiento, con un valor actual calculado en $260.28 millones con un valor futuro más probable de $1,332.11 Millones, siendo el crecimiento más probable del 512%.

**Referencias bibliográficas**

Justificación:

Grupo BIT (2018). 2022, 30 de Octubre “Todo Lo Que Debes Saber Sobre El Canal Tradicional.” Grupo BIT, 2018. Sitio web:

<https://www.grupobit.net/aprenda-con-grupo-bit/>

ISCAM:

ISCAM. (2022). Mercados. 2023, 15 de Abril, de ISCAM. Sitio web: <https://www.iscam.com/mercados/>

Datos Geoestadísticos:

INEGI. (2022). Marco Geoestadístico. 2023, 15 de Abril, de INEGI.ORG. Sitio web: <https://www.inegi.org.mx/temas/mg/#Descargas>

Geopandas:

Geopandas. (2022). Geocoding. 2023, 11 de Mayo, de Geopandas.org. Sitio web: <https://geopandas.org/en/stable/docs/user_guide/geocoding.html>

Folium:

Folium. (2023). Python data, leaflet.js maps. 2023, 11 de Mayo, de Python visualization. Sitio web: <https://python-visualization.github.io/folium/>

Filtros de pandas:

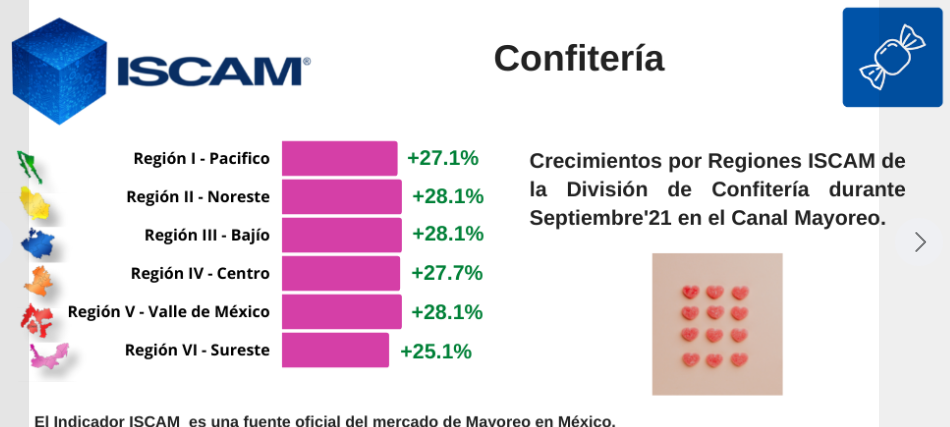
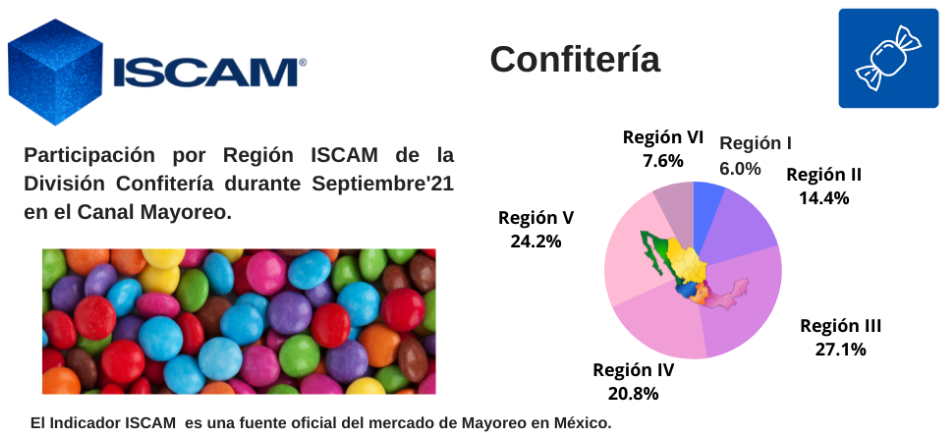
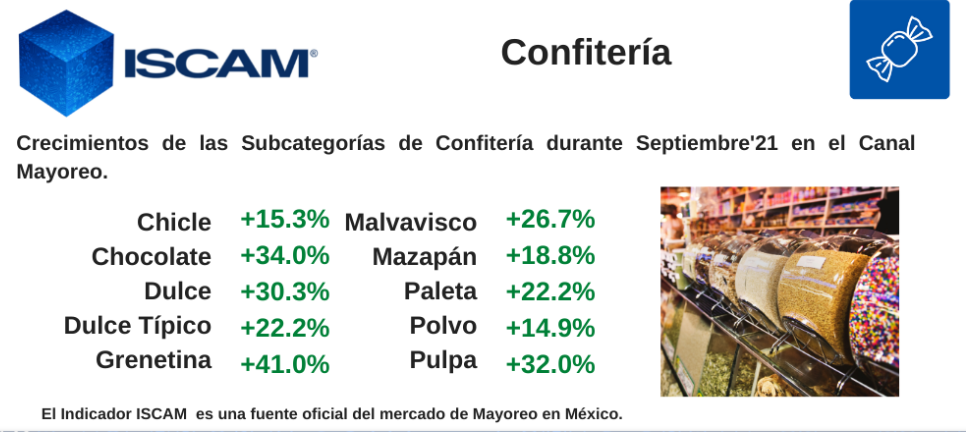
Pandas. (2023). pandas.DataFrame.filter. 2023, 11 de Mayo, de Pandas.pydata.org. Sitio web: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.filter.html>

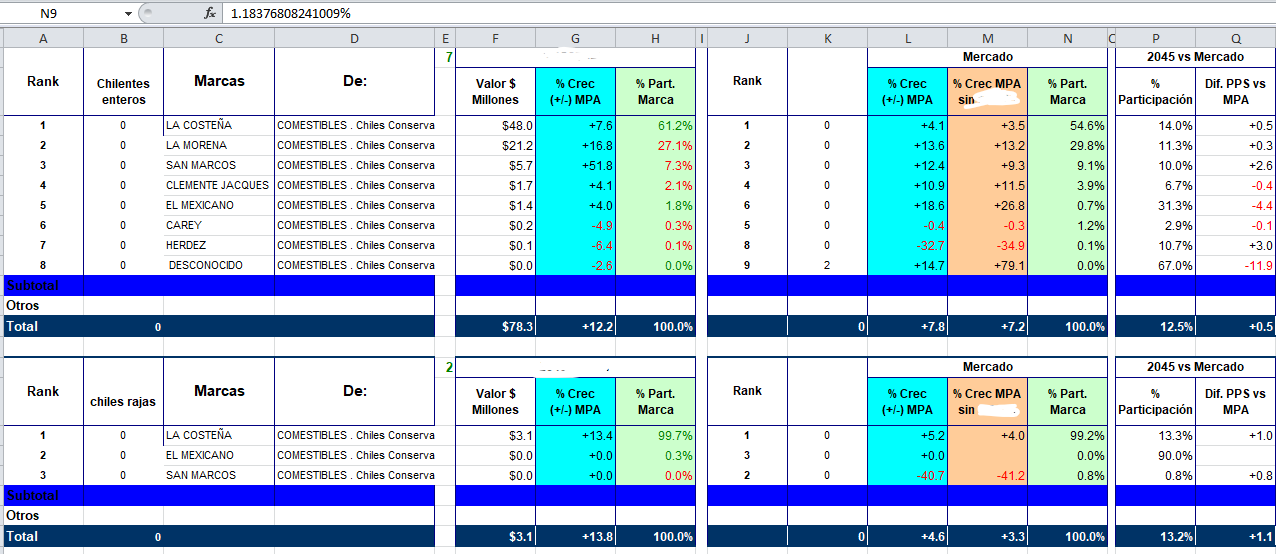
Documentación de Modelo Montecarlo:

Risk. (2023). DecisionTools 8.3. 2023, 11 de Mayo, de Palisade. Sitio web: <https://help.palisade.com/v8_3/en/Home.htm>

1. **Anexos**



**** **** 

****