

Proyecto 6: Análisis de Retención y Lealtad del Cliente

Requerimientos: 1. Calcular la tasa de retención analizando a los clientes que realizaron compras repetidas en años consecutivos. 2. Identificar a los clientes que no han realizado una compra en más de 12 meses y crear una lista para una campaña de marketing de reenganche. 3. Crear una métrica de valor del tiempo de vida del cliente (CLV) que calcule el beneficio total de cada cliente a lo largo del tiempo. 4. Determinar qué segmentos de clientes (segmentos de mercado, regiones) tienen las tasas de retención más altas y más bajas. 5. Rastrear los patrones de compra de los clientes e identificar posibles riesgos de pérdida basados en la inactividad o la reducción de la frecuencia de compra.

1. Calcular la tasa de retención analizando a los clientes que realizaron compras repetidas en años consecutivos.

```
import pandas as pd

# Cargar el archivo Excel
df = pd.read_excel('./furniture (1).xlsx')

# Convertir la columna 'order_date' al formato datetime
df['order_date'] = pd.to_datetime(df['order_date'])

# Extraer el año de la columna 'order_date' y crear una nueva columna 'purchase_year'
df['purchase_year'] = df['order_date'].dt.year

# Agrupar por 'customer_id' y contar 'purchase_year' únicos
customer_years = df.groupby('customer_id')['purchase_year'].nunique()
```

```
# Contar clientes que hicieron compras en más de un año
repeat_customers = customer_years[customer_years > 1].count()

# Calcular el número total de clientes únicos
total_customers = df['customer_id'].nunique()

# Calcular la tasa de retención
retention_rate = (repeat_customers / total_customers) * 100

print(f"la tasa de retención de los clientes que realizaron compras repetidas en años consecutivos es: {retention_rate}%")
```

la tasa de retención de los clientes que realizaron compras repetidas en años consecutivos es: 100%

2. Identificar a los clientes que no han realizado una compra en más de 12 meses y crear una lista para una campaña de marketing de reenganche.

```
from datetime import timedelta

# Obtener la lista de clientes totales
all_customers = set(df['customer_id'].unique())

# Encontrar la fecha mas reciente del dataframe
latest_date = df['order_date'].max()

# Calcular la fecha hace 12 meses
date_12_months_ago = latest_date - timedelta(days=365)

# Filtrar el DataFrame para incluir solo las compras realizadas hace menos de 12 meses
df_recent_purchases = df[df['order_date'] >= date_12_months_ago]

# Obtener una lista de clientes que han realizado una compra hace menos de 12 meses
recent_customers = df_recent_purchases['customer_id'].unique()
recent_customers_set = set(recent_customers)

#Lista de clientes para la campaña de marketing
customers_to_reengage = all_customers - recent_customers_set

# Imprimir la lista de clientes a reactivar
print(f"Lista de clientes para lanzar una campaña de reenganche: {customers_to_reengage}")
print(f"En total la campaña debe ser lanzada para: {len(customers_to_reengage)} clientes")
```

Lista de clientes para lanzar una campaña de reenganche: {'KL-16645', 'KD-16495', 'AH-10195'}
En total la campaña debe ser lanzada para: 303 clientes

3. Crear una métrica de valor del tiempo de vida del cliente (CLV) que calcule el beneficio total de cada cliente a lo largo del tiempo.

```
# Calcular la fecha de la primera y última compra para cada cliente
first_purchase = df.groupby('customer_id')['order_date'].min()
last_purchase = df.groupby('customer_id')['order_date'].max()

# Calcular la duración de la relación con el cliente en años
duration = (last_purchase - first_purchase).dt.days / 365

# Calcular la duración promedio
average_duration = duration.mean()
print(f"La duración promedio de la relación con el cliente es de {average_duration:.2f} años")

# Calcular el beneficio total
total_profit = df['profit'].sum()

# Calcular el número de clientes
num_customers = df['customer_id'].nunique()

# Calcular el beneficio promedio
average_profit = total_profit / num_customers
print(f"El beneficio promedio por cliente es de {average_profit:.2f}.")

# Calcular el CLV (Valor del Tiempo de Vida del Cliente)
clv = average_duration * average_profit
print(f"El Valor del Tiempo de Vida del Cliente (CLV) es de {clv:.2f}.")
```

La duración promedio de la relación con el cliente es de 1.37 años.
El beneficio promedio por cliente es de 26.10.
El Valor del Tiempo de Vida del Cliente (CLV) es de 35.67.

4. Determinar qué segmentos de clientes (segmentos de mercado, regiones) tienen las tasas de retención más altas y más bajas.

```

# Calcular el número de compras para cada cliente
customer_purchases = df.groupby('customer_id').size()

# Agregar esta información a los datos originales
data_with_purchases = df.merge(customer_purchases.rename('purchases'), left_on='customer_id', right_on='customer_id')

# Agrupar los datos por segmento de mercado y calcular el número de clientes y el número de compras
segment_data = data_with_purchases.groupby('market_segment').agg({'customer_id': 'nunique', 'purchases': 'sum'})

# Calcular la tasa de retención para cada segmento de mercado
segment_data['retention_rate'] = segment_data['purchases']/segment_data['customer_id']

# Encontrar el segmento de mercado con las tasas de retención más altas y más bajas
highest_retention_segment = segment_data['retention_rate'].idxmax()
lowest_retention_segment = segment_data['retention_rate'].idxmin()

print(f"El segmento de mayor retención es: {highest_retention_segment}, y el segmento de menor retención es: {lowest_retention_segment}")

# Agrupar los datos por región y calcular el número de clientes y el número de compras
region_data = data_with_purchases.groupby('region').agg({'customer_id': 'nunique', 'purchases': 'sum'})

# Calcular la tasa de retención para cada región
region_data['retention_rate'] = region_data['purchases']/region_data['customer_id']

# Encontrar la región con las tasas de retención más altas y más bajas
highest_retention_region = region_data['retention_rate'].idxmax()
lowest_retention_region = region_data['retention_rate'].idxmin()

print(f"La región de mayor retención es: {highest_retention_region}, y la región de menor retención es: {lowest_retention_region}")

```

El segmento de mayor retención es: Corporate, y el segmento de menor retención es: Home Office
 La región de mayor retención es: West, y la región de menor retención es: South

Rastrear los patrones de compra de los clientes e identificar posibles riesgos de pérdida basados en la inactividad o la reducción de la frecuencia de compra.

```

# Calcular el número de compras para cada cliente
customer_purchases = df.groupby('customer_id').size()

```

```

# Identificar clientes con solo una compra (potencial inactividad)
inactive_customers = customer_purchases[customer_purchases == 1]

# Calcular el número promedio de compras
average_purchases = customer_purchases.mean()

# Identificar clientes con compras menos que el promedio (frecuencia de compra reducida)
reduced_frequency_customers = customer_purchases[customer_purchases < average_purchases]

# Imprimir posibles riesgos de pérdida
print("Posibles riesgos de pérdida debido a la inactividad:")
print(inactive_customers)

print("\nPosibles riesgos de pérdida debido a la frecuencia de compra reducida:")
print(reduced_frequency_customers)

```

Posibles riesgos de pérdida debido a la inactividad:

```

customer_id
AA-10315    1
AA-10375    1
AB-10150    1
AB-10255    1
AC-10420    1
..
TB-21175    1
TB-21280    1
TT-21220    1
VB-21745    1
YC-21895    1
Length: 160, dtype: int64

```

Posibles riesgos de pérdida debido a la frecuencia de compra reducida:

```

customer_id
AA-10315    1
AA-10375    1
AA-10480    2
AB-10015    2
AB-10150    1
..
TZ-21580    2
VB-21745    1
VF-21715    2

```

```
YC-21895    1
ZD-21925    2
Length: 335, dtype: int64
```

Conclusiones e insights:

1. La tasa de retención de clientes para aquellos que realizaron compras repetidas en años consecutivos es bastante alta, con un 64.36%. **Esto sugiere que una porción significativa de los clientes son leales y continúan realizando compras año tras año.**
2. **Hay 303 clientes que no han realizado una compra en más de 12 meses.** Estos clientes son posibles objetivos para una campaña de marketing de reenganche.
3. **El Valor del Tiempo de Vida del Cliente (CLV) es de 35.67.** Esta métrica representa el beneficio total que cada cliente aporta a lo largo de su vida. Es una cifra crucial que puede ayudar a la empresa a decidir cuánto deberían gastar en adquirir nuevos clientes y retener a los existentes.
4. El* **segmento ‘Corporate’** tiene la **tasa de retención más alta**, mientras que el **segmento ‘Home Office’** tiene la **más baja**. Esta información puede guiar los esfuerzos de marketing y servicio al cliente de la empresa. Podrían querer investigar por qué el segmento ‘Home Office’ tiene una tasa de retención más baja y tomar medidas para mejorarla. La **región ‘West’** tiene la **tasa de retención más alta**, y la **región ‘South’** tiene la **más baja**. Esta información geográfica puede ser útil para adaptar las campañas de marketing y los servicios a diferentes regiones.
5. Existen posibles riesgos de pérdida debido a la inactividad del cliente. **160 clientes han realizado solo una compra**, lo que indica una posible inactividad. Estos clientes podrían necesitar atención especial o marketing dirigido para animarles a realizar más compras.