**Instrucciones:**

La forma en que se lleguen a los resultados no va a ser relevante, el objetivo principal es poder responder las preguntas. Se pueden utilizar tablas dinámicas en excel, Tableau, Python, R o la herramienta/software con la que se sientan más cómodos. Si quieren realizar cambios adicionales a las tablas para facilitar el análisis también es válido y si son útiles los cambios para el análisis se tomará en cuenta como algo positivo.

El documento Base\_Nielsen\_Prueba.xlsx tiene la información por formato y región de las ventas en monto y en unidades de los autoservicios. Hay 6 formatos: AUTO 1, AUTO 2, AUTO 3, AUTO 4, AUTO 5 y el RESTO. Los primeros 5 son los formatos pertenecientes a grupo Walmart México y el RESTO engloba los demás autoservicios (competencia)

La hoja Extracción Áreas tiene las ventas de los productos de marca comercial que se venden a nivel autoservicio y la hoja Extracción áreas MP tiene las ventas de productos de marca propia, éstos no tienen distinción desde el campo UPC hasta DESCRIPCIÓN (el valor es CONTROLLED LABEL y así se puede distinguir)

CONTROLLED LABEL son productos de la marca propia del autoservicio (incluyendo las marcas de la competencia) correspondiente, por ejemplo, en Walmart existen marcas como Aurrerá o Great Value. Como la información de estos productos es confidencial, no se puede saber la cantidad de productos englobados, precios individuales ni presentaciones. Considere estos puntos para no llegar a conclusiones erróneas en el análisis.

Las descripciones de las variables relevantes son las siguientes:

* Descripción: El nombre del ítem (Máxima granularidad de la base)
* Variable: Ventas en Valor en miles de pesos y las Ventas en Unidades en miles de unidades
* Mercado: Formato y Región unidos en una variable
* PMM’AA: Estas variables corresponden al mes y año, por ejemplo, P05’18 es mayo 2018

**Evaluación limpieza de datos y análisis**

**Parte 1**

El documento tiene 2 hojas, para ambas hojas se tienen que realizar las siguientes acciones:

1. Unir las 2 bases y poner la nueva tabla con toda la información en una nueva hoja que se llame “Base\_Nielsen”

Se concatenaron las bases de forma horizontal con python

1. Separar de la variable *Mercado* el Formato y la Región (La región son los últimos dos caracteres)

Usando una función lambda en python separando las variables con el espacio.

1. Una vez separados los formatos de las regiones cambie los valores de las regiones a los siguientes valores:

| A1 | PACIFICO |
| --- | --- |
| A2 | NORTE |
| A3 | OCCIDENTE |
| A4 | CENTRO |
| A5 | VDM |
| A6 | SURESTE |

Usando una función loc en python en la nueva columna de “Regiones”.

**Bonus:** Esta base de datos no tiene la estructura óptima para ser utilizada en herramientas de BI, identifique que procesos de limpieza se pueden aplicar para que sea más fácil utilizar la base de datos y realícenlos.

Se identificaron las columnas específicas de fecha, cambiandolas en formato %YYYY-%MM y posteriormente se utilizó la función melt para desagrupar los valores a partir de marca, rollos, presentación unificada, descripción, tier, variable, formato y región. la métrica faltante se denominó cantidad comprada y fue definida a partir de las fechas identificadas.

**Parte 2**

El objetivo del ejercicio es identificar en donde estamos perdiendo ventas en monto, comparando 2 períodos que sean comparables. Responde las siguientes preguntas utilizando la tabla de Base\_Nielsen:

1. ¿Qué conjuntos de periodos de tiempo utilizarías para comparar las ventas y por qué?

utilizaría los periodos de enero a abril de 2019 y de enero a abril de 2020. Esto debido a que se atravesó el incidente de la pandemia. Considerando que hubieron compras de pánico de productos de limpieza, podría haber un cambio significativo en estos periodos.

1. En base a los periodos elegidos, hacer un comparativa de las ventas en monto y responda la siguiente pregunta: ¿Cuáles son los 2 formatos que más ventas en monto han perdido y cuánto?

Considerando la moneda de cambio a pesos mexicanos, los dos formatos con más pérdidas son AUTO2 y AUTO3 con -$34,651.25MXN y -$26,147.125MXN respectivamente

1. ¿Cuánto Market Share han perdido/ganado todos los formatos? (Market Share = Suma de ventas en monto del formato / Suma de ventas en monto totales).

| Formato | Cantidad comprada | Market Share (%) |
| --- | --- | --- |
| AUTO1 | $2,832,737,000,000.00 | 10.00% |
| AUTO2 | $8,034,869,000,000.00 | 29.00% |
| AUTO3 | $4,035,814,000,000.00 | 14.00% |
| AUTO4 | $185,880,800,000.00 | 10.00% |
| AUTO5 | $1,337,619,000,000.00 | 50.00% |
| RESTO | $11,695,760,000,000.00 | 42.00% |

1. Haz una matriz con los formatos en los renglones y las regiones en las columnas y los valores que sean las ventas en monto. ¿De qué dimensiones queda la matriz?

Se tiene una matriz de 6x6 las cuales corresponden a las 6 regiones y los 6 formatos.

1. ¿Cuáles son las 5 combinaciones de Formato/Región en donde se perdieron más ventas en monto? (usar matriz del inciso anterior para responder la pregunta)

| **Region** | **CENTRO** | **NORTE** | **OCCIDENTE** | **PACIFICO** | **SURESTE** | **VDM** |
| --- | --- | --- | --- | --- | --- | --- |
| **Formato** |  |  |  |  |  |  |
| **AUTO1** | $788,401.13 | $392,555.50 | $605,040.00 | $232,239.88 | $711,658.63 | $102,841.88 |
| **AUTO2** | $1,508,966.50 | $1,215,057.25 | $1,353,149.38 | $632,740.63 | $1,339,634.75 | $1,985,320.63 |
| **AUTO3** | $531,036.63 | $653,104.75 | $754,643.50 | $445,727.25 | $451,315.25 | $1,199,986.25 |
| **AUTO4** | $27,268.00 | $5,053.75 | $22,422.75 | NaN | $14,510.38 | $116,625.88 |
| **AUTO5** | $128,516.13 | $408,029.88 | $292,470.75 | $41,237.38 | $45,521.25 | $421,844.00 |
| **RESTO** | 1143925.25 | $3,365,357.13 | $1,061,060.00 | $2,301,561.13 | $2,220,741.13 | $1,603,114.38 |

Para estimar las 5 peores combinaciones de formato y region se determinó la siguiente gráfica

| Formato | Region | Fecha\_inicial | Cantidad comprada inicial | Fecha\_final | Cantidad comprada final | diferencia |
| --- | --- | --- | --- | --- | --- | --- |
| AUTO3 | VDM | 2018-05 | 47100.25 | 2020-04 | 37945.625 | -9154.625 |
| AUTO2 | VDM | 2018-05 | $79,974.38 | 2020-04 | $71,558.50 | -8415.875 |
| AUTO2 | OCCIDENTE | 2018-05 | $52,882.25 | 2020-04 | $45,050.88 | -7831.375 |
| AUTO2 | CENTRO | 2018-05 | $58,625.75 | 2020-04 | $51,134.63 | -7491.125 |
| AUTO2 | SURESTE | 2018-05 | $48,709.88 | 2020-04 | $43,663.38 | -5046.5 |

**Parte 3**

Utilizando el Formato/Región que más ventas en monto ha perdido y los periodos elegidos en la Parte 2, responda las siguientes preguntas:

1. ¿Cuáles son los productos con más ventas en monto en el periodo más reciente? Top 10

Descartando los productos “controlled label” los 10 productos mas vendidos son los siguientes:

| DESCRIPCIÓN | Cantidad comprada |
| --- | --- |
| KBB SUAVELASTIC MAX ET 5 JUMBO 60PZAS 75019434... | $35,627.38 |
| KLEEN BEBE SUAVELASTIC MAX PANAL EXTRA JUMBO T... | $31,926.13 |
| HUGGIES ULTRACONFORT PANAL ETAPA 6 NINO BSA 60... | $27,874.38 |
| KLEENBEBE MOVILASTIC E6 XJ 40 PZAS 7506425603251 | $26,140.75 |
| CHICOLASTIC KIDDIES ANTIFUGAS ETAPA 5 BOLSA 40... | $26,075.13 |
| HUGGIES ULTRACONFORT PANAL ETAPA 6 NINA BSA 60... | $24,970.38 |
| HUGGIES ALL AROUND PANAL ETAPA 6 UNISEX BOLSA ... | $23,935.63 |
| CHICOLASTIC KIDDIES ANTIFUGAS ET6 BOLSA 40 PAN... | $22,618.88 |
| KBB SUAVELASTIC MAX ET 4 GRANDE 60PZAS 7501943... | $21,431.13 |

1. ¿Cuál es el precio de los productos que más vendemos? y ¿Cómo calculaste el precio?

Se calculo dividiendo la base en dos partes, una cuya “Variable” fuera igual a “ventas en valor (in 000 pesos) la cual es el precio de los productos y la segunda parte cuya “Variable” fuese igual a “ventas en unidades (in 000)”. Se descartaron las de controlled label dado que pueden causar un sesgo en el analisis.

Posteriormente se dividió la cantidad comprada de “ventas en valor (in 000 pesos)” entre la cantidad comprada de “venta en unidades (in 000). Si la moneda de la tabla son los pesos mexicanos, el top 3 es el siguiente

| DESCRIPCION | Cantidad comprada\_precio | Cantidad comprada\_unidad | PRECIO POR PIEZA |
| --- | --- | --- | --- |
| KBB SUAVELASTIC MAX ET 5 JUMBO 60PZAS 75019434... | $15036.12 | $87.63 | $171.60 |
| CHICOLASTIC KIDDIES ANTIFUGAS ETAPA 5 BOLSA 40... | $11032.75 | $85.25 | $129.42 |
| KBB SUAVELASTIC MAX ET 5 JUMBO 60PZAS 75019434... | $9608.125 | $48.50 | $198.10 |

1. ¿Cuáles son los productos que más ventas en monto han perdido? (se necesita en absolutos y el % de Decrecimiento).

| DESCRIPCIÓN | ventas 05-2018 | ventas 04-202 | variable absoluta | variable porcentual |
| --- | --- | --- | --- | --- |
| KLEEN BEBE SUAVELASTIC MAX PANAL EXTRA JUMBO T... | $42,698.88 | $31,926.13 | -$10,772.75 | -25.23 |
| CHICOLASTIC KIDDIES JUMBO T5 BSA 80PZ 00131170... | $16,920.63 | $6,976.13 | -$9,944.50 | -58.77 |
| KBB SUAVELASTIC MAX ET 4 GRANDE 60PZAS 7501943... | $31,055.38 | $21,431.13 | -$9,624.25 | -30.99 |
| KBB SUAVELASTIC MAX ET 5 JUMBO 60PZAS 75019434... | $44,242.25 | $35,627.38 | -$8,614.88 | -19.47 |
| HUGGIES ULTRACONFORT PANAL ETAPA 4 NINA BSA 60... | $20,090.63 | $12,976.13 | -$7,114.50 | -35.41 |
| KBB COMODISEC E5 JUMBO BOLSA 60 PZAS 750642560... | $19,601.00 | $12,609.38 | -$6,991.63 | -35.67 |
| HUGGIES ULTRACONFORT PANAL ETAPA 4 NINO BSA 60... | $16,087.25 | $9,566.50 | -$6,520.75 | -40.53 |
| CHICOLASTIC KIDDIES PANALES GRANDE ET4 NINO BS... | $9,069.50 | $3,276.63 | -$5,792.88 | -63.87 |
| HUGGIES SUPREME PANAL ETAPA 1 38PZA 7501943416864 | $9,861.00 | $4,627.00 | -$5,234.00 | -53.08 |
| CHICOLASTIC BB TIPS ETAPA 5 BOLSA 62 PANALES N... | $21,006.38 | $15,850.13 | -$5,156.25 | -24.55 |

1. ¿Qué fabricantes han perdido más ventas? Top 5 (se necesita en absolutos y el % de Decrecimiento).

| FABRICANTE | ventas\_inicio | ventas\_fin | variacion\_abs | variacion\_pct |
| --- | --- | --- | --- | --- |
| MABESA | $280,580.88 | $240,580.75 | -$40,000.13 | -14.256184 |
| OTHERS FABRICANTE UNIF. | $29,757.13 | $19,449.63 | -$10,307.50 | -34.638763 |
| KIMBERLY-CLARK | $610,877.75 | $606,106.00 | -$4,771.75 | -0.78113 |
| THE HONEST CO. | $79.50 | $6.25 | -$73.25 | -92.138365 |
| LAMBI | $2,684.25 | $2,683.00 | -$1.25 | -0.046568 |

1. ¿Qué marcas han perdido más ventas? Top 5 (se necesita en absolutos y el % de Decrecimiento).

| MARCA | ventas\_inicio | ventas\_fin | variacion\_abs | variacion\_pct |
| --- | --- | --- | --- | --- |
| CHICOLASTIC | $250,440.63 | $216,534.75 | -$33,905.88 | -13.54 |
| HUGGIES | $272,686.50 | $257,668.38 | -$15,018.13 | -5.51 |
| OTHERS MARCA UNIF. | $29,761.50 | $19,449.63 | -$10,311.88 | -34.65 |
| BABY PANTS | $9,162.88 | $4,867.38 | -$4,295.50 | -46.88 |
| BIOBABY | $19,829.00 | $18,286.88 | -$1,542.13 | -7.78 |

1. ¿Cuáles de todos los productos que más venta en monto perdemos son los más relevantes? Top 5

| DESCRIPCION | ventas\_inicio | ventas\_fin | variable absoluta | variable porcentual | participacion\_inicio |
| --- | --- | --- | --- | --- | --- |
| KBB SUAVELASTIC MAX ET 5 JUMBO 60PZAS 75019434... | $44,242.25 | $35,627.38 | -$8,614.88 | -19.472054 | 0.043433 |
| KLEEN BEBE SUAVELASTIC MAX PANAL EXTRA JUMBO T... | $42,698.88 | $31,926.13 | -$10,772.75 | -25.229587 | 0.041918 |
| KBB SUAVELASTIC MAX ET 4 GRANDE 60PZAS 7501943... | $31,055.38 | $21,431.13 | -$9,624.25 | -30.99061 | 0.030488 |
| HUGGIES ULTRACONFORT PANAL ETAPA 6 NINO BSA 60... | $30,136.63 | $27,874.38 | -$2,262.25 | -7.506647 | 0.029586 |
| HUGGIES ULTRACONFORT PANAL ETAPA 6 NINA BSA 60... | $27,456.63 | $24,970.38 | -$2,486.25 | -9.055192 | 0.026955 |

**Parte 4**

Responde a las siguientes conceptuales:

1. ¿Podemos incrementar ventas, pero perder Market Share? (Por ejemplo, que el Formato 1 incremente en 10 MDP vs periodo anterior, pero que pierda 1% de Market Share vs periodo anterior)

Si si aumenta en general el resto de áreas de formato y región las ventas de una forma mas proporcionada en comparación al área que estamos analizando

1. Si encuentras 2 productos de la misma marca, pero con diferentes conteos, (por ejemplo, six pack de Victoria vs 12 pack de Victoria), ¿Cómo podrías decidir cuál te conviene comprar?

Revisando el costo por litro de un six pack de victoria vs el 12 de victoria:

**Parte 5**

Haz un análisis completo de la categoría, el análisis deberá contener los siguientes puntos:

1. Overview general de la categoría, con el desempeño de los formatos en el mercado.
2. Identificar el formato de Walmart con mayor pérdida y profundizar sobre este.
3. Profundización formato: Se tiene que llegar hasta nivel item (producto), mostrando de lo general a lo particular como se identificó el insight en los productos seleccionados en el análisis. (Importante: Análisis que no lleguen hasta este nivel de granularidad, será descartado del proceso).
4. Proponer una estrategia de recuperación sobre los productos encontrados en el inciso anterior.

Bonus : Usar fuentes externas que complementen el análisis ( Ejemplo: INEGI).

**Parte 6**

Haz una presentación en donde muestres los resultados de la **Parte 5**. Debes de ser capaz de explicar el análisis obtenido en el inciso anterior. La presentación se debe de hacer en power point y debe ser enviada en la fecha asignada.

**Importante : Cuidar formato, uso de gráficos claros que expliquen los insights, evitar demasiado uso de texto o tablas.**  **Presentación que no cumpla con estos criterios, será descartada.**

**Parte 7**

Presentación Final, si la presentación cumple con los requisitos solicitados, se tendrá que realizar una presentación de 30 minutos donde expliques los resultados obtenidos. (Se te notificará cuando la presentas)

**Evaluación de Herramientas de BI**

**Parte 8**

Esta parte corresponde al conocimiento técnico sobre herramientas de BI (Tableau Prep o Power BI). Las preguntas no van enfocadas hacia una herramienta/software en específico.

1. Se necesita hacer una conexión a una fuente de datos con el siguiente Query:

Select \* from <tabla>

¿Cuáles son las mejores prácticas para realizar este tipo de conexiones?

Se recomienda realizar siempre un commit al finalizar la consulta, además de delimitar la consulta a partir de las necesidades de negocio, ya sea por año, por area, por region, etc. Ejemplo: select \* from <table> WHERE *CONDICION.*

El caso contrario de consultar todos los datos sin una delimitación adecuada, alentará la ejecución de dashbard

1. El equipo de desarrollo en Backend comparte los datos de conexión, esta tiene una tabla que une todas las extracciones hechas por ellos. ¿Que se validaría dentro de la tabla para que fuera funcional para una herramienta de visualización?.

* Que los datos de fecha se encuentren en el formato adecuado.
* Verificar que los join se hagan mediante las primary keys adecuadas.
* Tener siempre un ejemplo de datos reales para realizar el qa de la base de datos creada en el backend.

1. Se recibe un desarrollo de parte de un proveedor externo. Se busca optimizar este desarrollo ya que el usuario final reporta lentitud en algunos objetos. ¿Cuál definirías como la causa principal de este problema? Selecciona una opción y comenta porque descartaste las otras
2. La tabla contiene más de 10 tablas dentro del modelo de Front.
3. La conexión es en vivo.
4. La mayoría de los campos en Front son formulas dentro del tablero.

La opción c es la que presentaría mas lentitud, ya que esta consumiría parte de la ram del equipo, además de que cada vez que ejecute el dashboard recalcula los valores de manera constante (al menos esto ocurre en tableau y looker studio). En cuanto a la opción a puede no alentar el dashboard si tiene los campos óptimos y la opción b no siempre alenta el dashboard si la conexión es nativa a la fuente de datos, como es el caso de las API.

1. El usuario requiere un desarrollo que muestre a nivel item un tablero, con varios años de historia. Esto provoca que el equipo Back desarrolle tablas de Millones de registros. ¿De qué forma se debería abordar el tema de la volumetría en las herramientas visuales?

Solo extraer las dimensiones necesarias para el análisis, tratar de agrupar los datos para agilizar la lectura de éstos. También evitar usar las dimensiones calculadas lo más posible para evitar alentar el dashboard y su performance. Evitar utilizar fuentes estáticas de información como archivos planos y optar por apis.

**Parte 9**

**Dashboard**

Utilizando Power BI diseñen un dashboard en donde se pueda ver la información más relevante de la base de datos de Excel adjuntado. El objetivo del dashboard es poder identificar de forma eficiente la información y poder ver a alto nivel las oportunidades generales.