**Lectura Previa de data analytics**

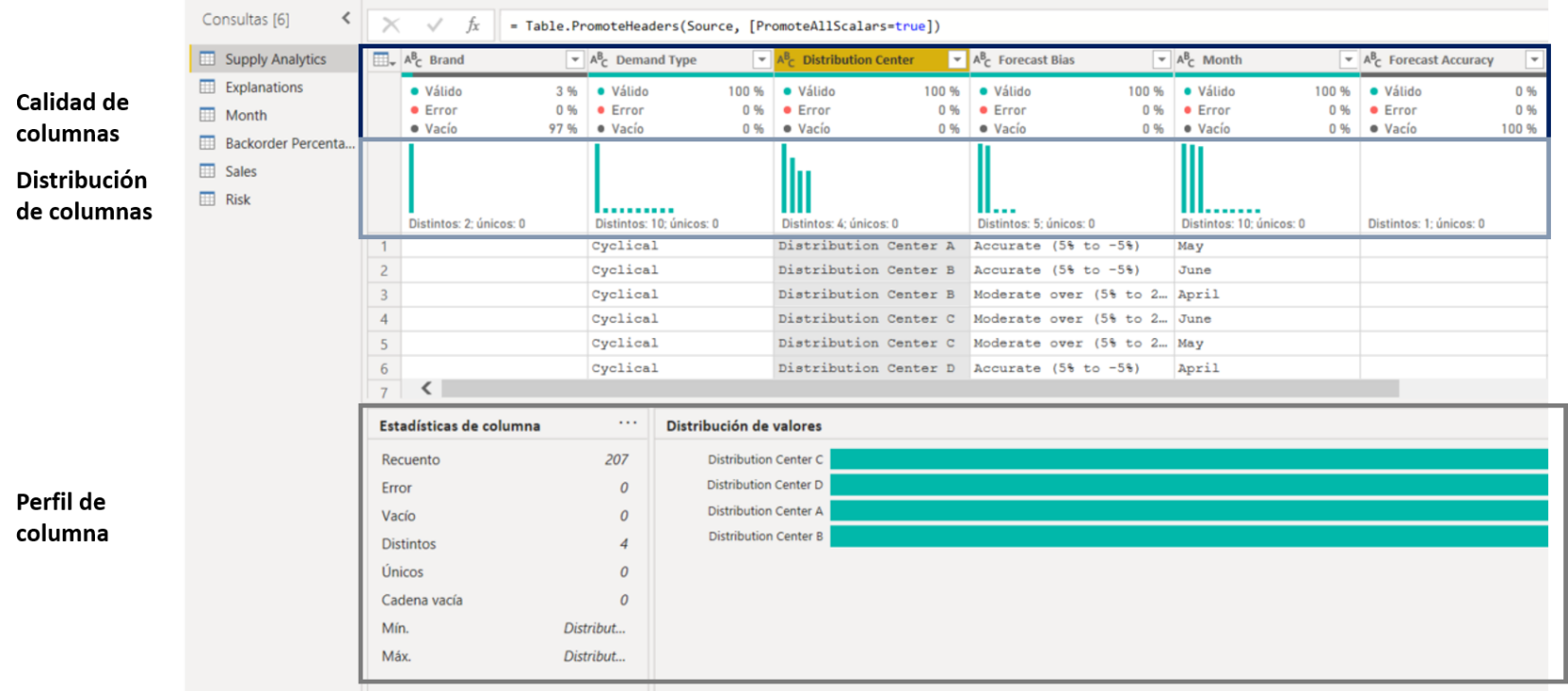
**Tarea 1: Introducción a analytics, estadísticos y Analize**

Power BI posee múltiples utilidades y herramientas del data analytics. A continuación, se describen algunas de las más relevantes:

* **Expresión de análisis de datos (**DAX): Es un lenguaje fácil e intuitivo, similar al empleado en Excel, que permite generar cálculos y métricas. Entre otras utilidades, permite calcular los estadísticos convencionales de distribución como media, moda, mediana, etc.
* **Integración con R y Python:** Power BI posee conectores de R y Python que permiten ejecutar scripts. De esta manera podemos utilizar todas las capacidades analíticas de unos de los lenguajes mas, utilizados para el análisis de datos.
* **Técnicas de segmentación en clústeres:**  Divide o combina diferentes conjuntos de datos en un grupo. A continuación, compara esos grupos para identificar una relación significativa entre los conjuntos de datos.
* **Series temporales**: permiten identificar anomalías en la distribución de los datos a lo largo del tiempo y realizar predicciones.
* **Función Analize o insights**: Los insights, representados por la función Analize, son un conjunto de análisis dentro de las gráficas tradicionales que permite el análisis en profundidad con funciones de IA. Dentro de los insights encontramos: **explicar el aumento/disminución** y **buscar en qué se diferencia esta distribución**.
* **Quick insights:** proporcionan un análisis automático de los datos en Power BI Service.
* **Influenciadores clave**: Muestran el conjunto de impulsores que contribuyen al acontecimiento de un determinado evento y la intensidad en su contribución.
* **Árbol de descomposición;** es un elemento visual altamente interactivo que permite descomponer una medida visualmente en diferentes dimensiones

**Tarea 2: Análisis de la calidad del dato e identificación de outliers**

Power BI permite **conocer la calidad de los datos** con la utilidad de[**Data Profiling, Quality y Distribution feature**](https://www.ayraanalytics.com.au/data-profiling-in-power-bi/#:~:text=During%20the%20process%20of%20data,desktop%20as%20a%20preview%20feature.). Esta utilidad permite, de una forma visual, explorar los datos y tener una idea de la composición. Obtiene distribuciones de columna, verifica cuántos errores hay en una columna determinada y cuántos valores únicos tiene, así como algunos otros estadísticos.



Por otro lado, además de la calidad de los datos conviene analizar si hay outliers o anomalías en nuestros datos. Un **outlier**es una observación que dista bastante del comportamiento del resto de datos, es decir, presenta unas características atípicas. Para la detección de oultiers podemos utilizar **gráficos de dispersión, box plots, histogramas**, etc. En Power BI también hay **gráficos avanzados en R** especialmente creados para el análisis de outliers. En este proyecto utilizaremos el[gráfico de outliers](https://appsource.microsoft.com/en-us/product/power-bi-visuals/WA104381449?tab=Overview&signInModalType=2&ctaType=1).

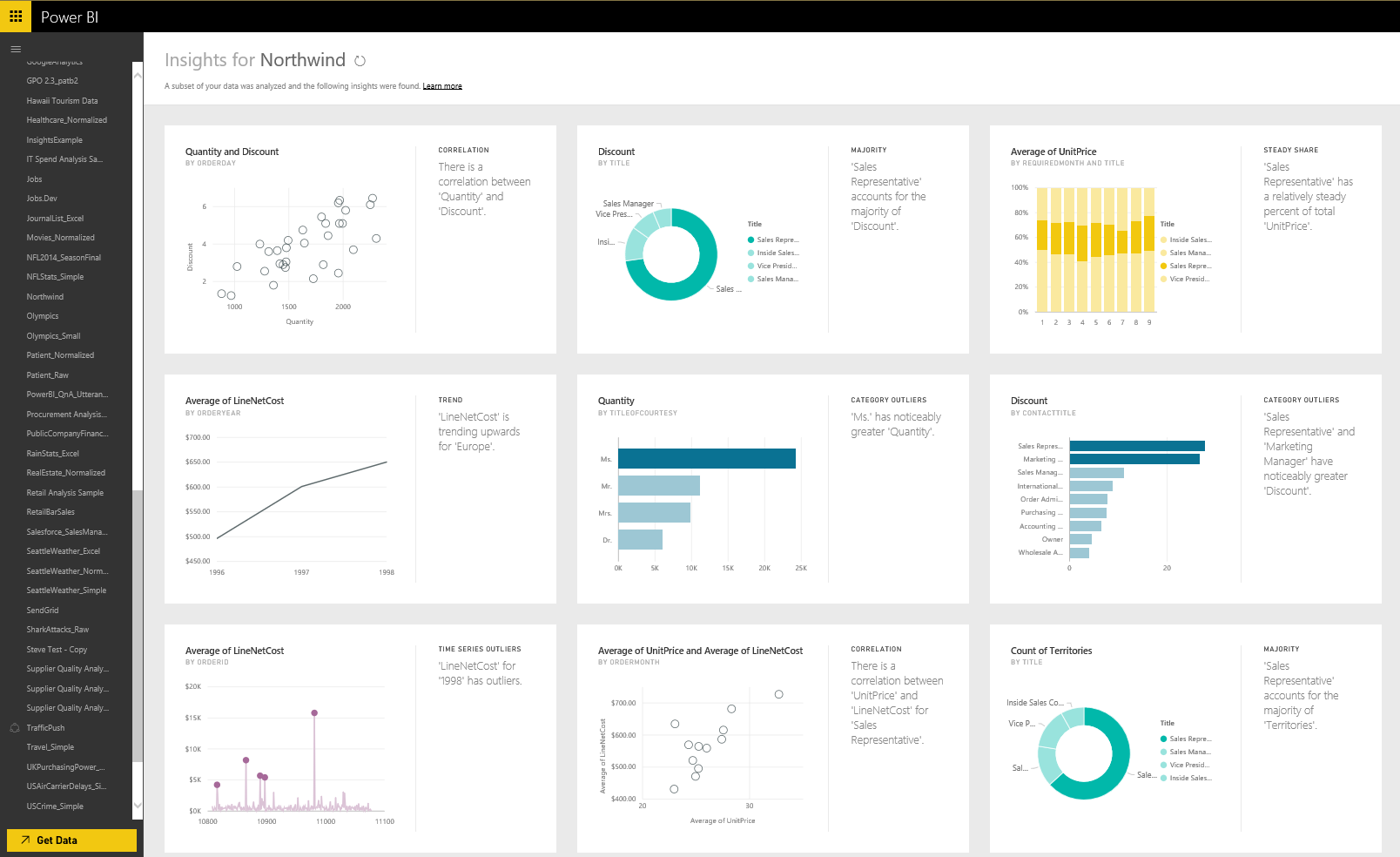
Hay muchos **estadísticos** **para detectar valores atípico**s. En este Visual, se utilizan cinco estadísticos de detección muy populares:

1. **Zscore**es un enfoque de desviación estándar.
2. **Método de Tuke**y, que utiliza un enfoque de rango intercuartílico (IQR)
3. **Método LOF** (Local Outlier Factor), que calcula la desviación de la densidad local de un punto de datos dado con respecto a sus vecinos. Considera las muestras que tienen una densidad sustancialmente más baja que sus vecinas como valores atípicos.
4. **Distancia de Cook**, que muestra la influencia de cada observación en los valores de respuesta ajustados.
5. Finalmente, el **método manual**le permite definir sus propios umbrales superior e inferior para los datos originales.

Además de esos métodos, contiene t**res tipos de gráficos** para visualizar los valores atípicos: **Gráfico de dispersión, Diagrama de caja** y **Gráfico de densidad.**

**Tarea 3: Quick insights**

Los **[quick insight](https://docs.microsoft.com/en-us/power-bi/create-reports/service-insights" \t "_blank)**s permiten obtener información relevante acerca de la distribución de los datos. Esta utilidad se encuentra en Power BI Service, por lo que es necesario publicar el informe en **Power BI Service**.



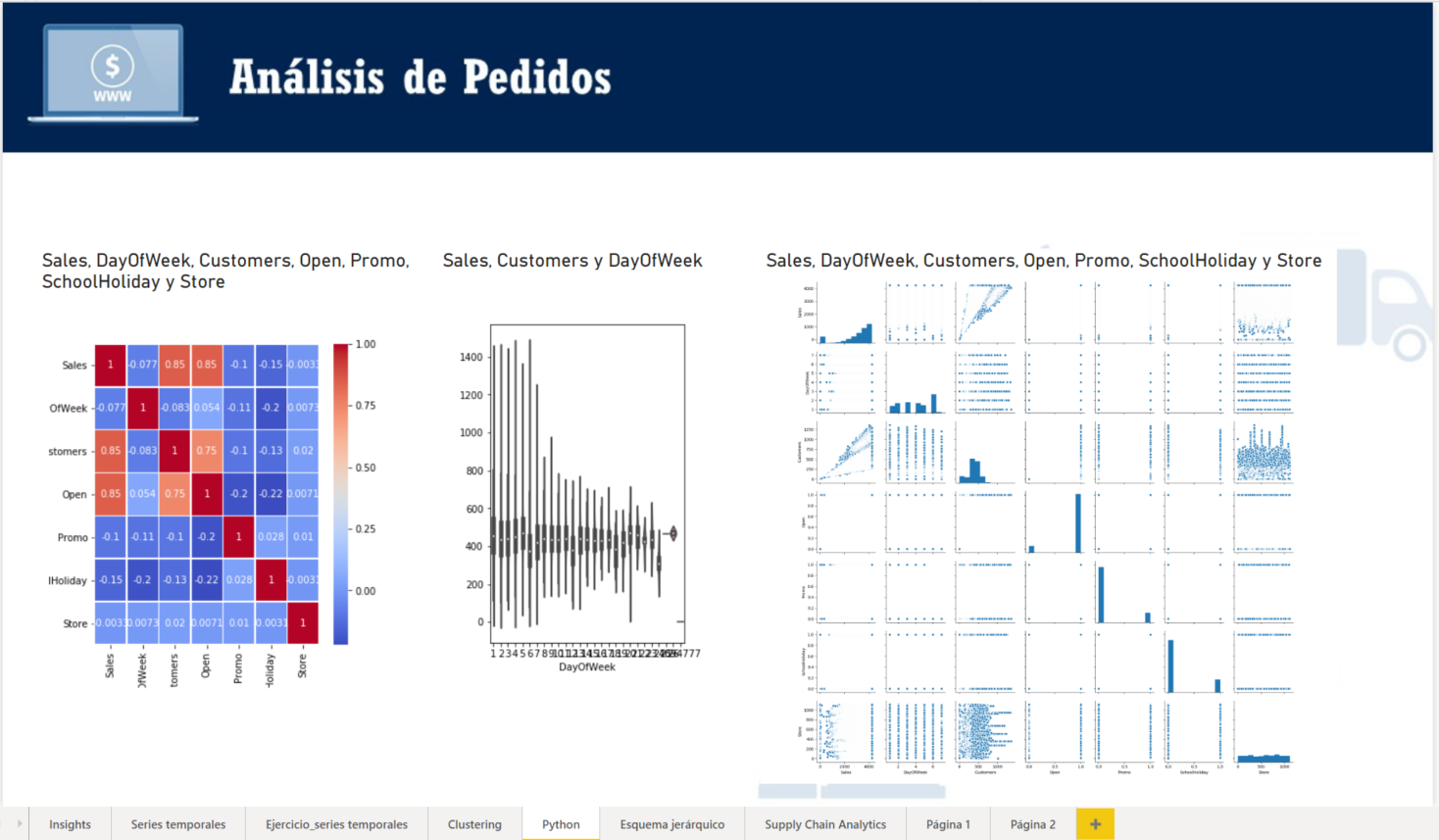
**Tarea 4: Técnicas de clustering y series temporales**

La **agrupación en clústers** permite identificar un grupo (cluster) de**datos que son similares entre sí**, pero diferentes al resto de los datos. Power BI permite la identificación automatizada de estos clústers, asigna de manera automática del numero de clústers y también permite la selección manual del mismo. Esta técnica es de utilidad para identificar grupos de cosas que se comportan de manera similar, como identificar grupos de clientes, por ejemplo.

Además de las técnicas de clustering, encontramos varias utilidades relacionadas con el análisis y predicción de series temporales. El **análisis de series temporales**se utiliza para analizar los datos a lo largo del tiempo. Se utilizan principalmente **para identificar anomalías, realizar predicciones, identificar tendencias**, etc. Hay dos utilidades relevantes en el análisis de series temporales. **El eje de reproducción**, que permite una visualización dinámica de la información. Y la **predicción de series temporales** con la utilidad de Forecasting.

**Tarea 5: Análisis avanzado con Python**

**Python**permite integrar numerosas utilidades y f**unciones avanzadas de analítica de datos**. Para ello, basta con instalar Python en nuestro equipo y los paquetes necesarios. Una vez configurado el Power BI para que utilice el entorno que queremos utilizar, basta con introducir el script de Python. En este proyecto se utilizará la librería de visualización [Seaborn](https://seaborn.pydata.org/examples/index.html" \t "_blank)para el análisis exploratorio de los datos, el análisis de correlaciones y la visualización de los datos.



**Tarea 6: Visuales de analítica avanzada**

Dentro de los visuales de analítica avanzada con **funcionalidades de Inteligencia Artificial (IA)** tenemos el esquema jerárquico y el identificador de influenciadores clave.

* **Influenciadores clave**: Muestran el conjunto de impulsores que contribuyen al acontecimiento de un determinado y la intensidad en su contribución. **Tiene dos tipos de análisis.** La opción de **Influenciadores clave** muestra los principales factores que contribuyen al valor de la métrica seleccionada. La opción de **Segmentos principales** muestra los segmentos principales que contribuyen al valor de la métrica seleccionada
* **Árbol de descomposición:**es un elemento visual altamente interactivo que permite descomponer una medida visualmente en diferentes dimensiones. Este elemento visual se puede utilizar para análisis exploratorios ad hoc para comprender los datos. O se puede utilizar para realizar un análisis de la causa raíz complementando a los influenciadores clave. Contiene la funcionalidad de IA conocida como Divisiones de IA, que encuentra los valores altos y bajos de los datos de forma automática.

Además de los elementos de visualización por defecto, podemos importar otros **gráficos de la AppStore**. Recomendamos echar un vistazo a las **gráficas de análisis avanzado de**[**R.**](https://community.powerbi.com/t5/R-Script-Showcase/bd-p/RVisuals)

