



**CURSO 2020-2021**

**MÁSTER EN BUSINESS INTELLIGENCE Y DATA SCIENCE**

**MODULO:**

**Business Intelligence**

**Nombre estudiante: Marc Faravelli Rodríguez**

**No.Matricula: 2047104**

**E-mail: marcfaravelli@gmail.com**

## Índice

<b>Fase 1 – Plan de proyecto.....</b>	<b>3</b>
<b>Fase 2 – Interpretación de datos con SQL.....</b>	<b>8</b>
<b>Fase 3 – Aplicando los conocimientos en Power BI.....</b>	<b>13</b>
<b>Fase 4 – Implementación / Visualización de datos en Power BI en un caso real.....</b>	<b>14</b>

## Fase 1 – Plan de proyecto

### Objetivo del proyecto

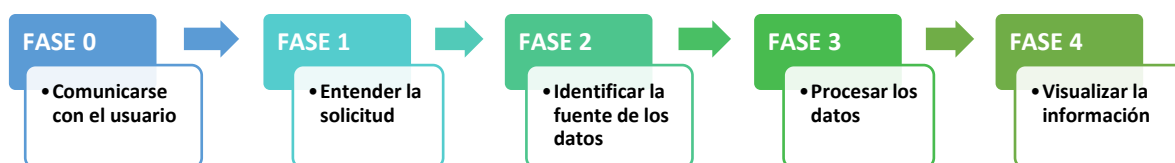
En esta fase preliminar, se realiza una verificación de la situación empresarial desde varios puntos de vista. En primer lugar, es necesario establecer cuáles son las necesidades analíticas y qué problemas se presentan en la empresa por su insatisfacción. Posteriormente, es necesario comprender qué herramientas de hardware y software ya están presentes en la empresa, tanto para adquirir conocimientos de sistemas operativos como para comprender el alcance actual de las tareas administrativas y de mantenimiento.

La Business Intelligence es un formidable aliado para organizar nuestra empresa de la forma más eficiente posible. Por ejemplo, gracias a informes y análisis, podemos establecer qué ROI (Return On Investment) se está produciendo, permitiendo así perfeccionar la estrategia de marketing en la que hemos invertido.

Además, al analizar en detalle los datos de nuestro negocio, BI puede identificar fácilmente qué áreas de nuestra empresa necesitamos mejorar, optimizando también las interacciones entre los diversos sectores mediante el análisis de datos y patrones recurrentes que, de lo contrario, estarían sin conexión. Esto significaría aumentar la eficiencia de nuestra empresa.

### Fases del proyecto

En el siguiente diagrama, se muestran las diferentes fases que componen el proyecto de Business Intelligence de la empresa Adventure Works Cycles Inc:



- **Fase 0: comunicarse con el usuario**

El analista de BI es ante todo capaz de interactuar con las figuras de cada departamento de la empresa, entendiendo su lenguaje y educándolos en el uso de una terminología única, que puede ser compartida por el resto de la empresa. De hecho, a menudo sucede que el mismo concepto se expresa en términos diferentes entre diferentes oficinas, o que se use el mismo término para conceptos absolutamente diferentes.

- **Fase 1: entender la solicitud**

Cuanto menos consolidado sea el uso de un vocabulario corporativo único, más complicado será comprender las necesidades del usuario. Además, el usuario, que conoce muy bien sus procesos, realiza breves solicitudes en las que omite detalles de fundamental importancia. Para él estos detalles son obvios, es superfluo explicitarlos. Si el usuario no está dispuesto a interactuar con los analistas de BI o la desorganización interna no lo permite, el trabajo se ve entorpecido y se completa con errores que el usuario deberá identificar y sugerir cómo corregir. Esto crea un rebote de responsabilidad que solo hará perder el tiempo a todos, elevando los costes de implementación. Por eso es fundamental desarrollar una terminología clara y compartida entre todos los miembros de la organización empresarial.

- **Fase 2: identificar la fuente de los datos**

Las fuentes de datos de la empresa son las más diversas. Los ejemplos más comunes son bases de datos, hojas de cálculo y documentos de texto, pero no hay límite para la imaginación que una empresa puede tener para almacenar información. Cuando un usuario realiza una solicitud de Business Intelligence, el analista de BI debe poder saber dónde se encuentran los datos que luego tendrá que volver a procesar. La gran mayoría de los datos se almacenan en las bases de datos, pero es necesario conocer la fuente exacta dentro de la base de datos para poder proceder con la extracción de los datos.

- **Fase 3: procesar los datos**

Los datos procesados se almacenarán en el data warehouse. Es una pieza importante detrás de todo proyecto de Business Intelligence: la calidad de todo el proyecto y la facilidad con la que se puede llevar a cabo la cuarta fase, que es la presentación visual de la información, depende de su calidad. Los dos pasos de reprocesar y almacenar los datos junto con la extracción forman lo que se llama ETL. El procesamiento de datos deberá seguir reglas comerciales precisas, que la empresa ha adoptado a lo largo de los años. El data warehouse tiene la ventaja de hacer explícitas estas reglas y mantenerlas documentadas, compartiendo su lógica dentro de la empresa.

Cuanto más anticuado y heterogéneo está el sistema de información de la empresa, más se requiere una fase sólida y estructurada de procesamiento de datos. Cada empresa tiene su propia selección personal de sistemas de gestión o ERP (Enterprise Resource Planning), que pueden variar tecnológicamente incluso dentro de la misma empresa. Es en estos casos que el trabajo de integración de datos es fundamental. El procesamiento de los datos también

consiste en una ampliación, mejora y fortalecimiento real de la información que pone a disposición el sistema de información de la empresa.

- **Fase 4: visualizar la información**

Los usuarios que utilizan Business Intelligence tienen acceso única y exclusivamente al producto final de esta fase. Es lo que comúnmente se denomina reporting, que es un conjunto de documentos que muestran gráficamente, en diversas formas y con distintos modos de acceso e interacción, datos e información generados por ellos.

Es extremadamente importante presentar los datos de la manera más correcta. Cuanto más intuitivo y atractivo sea el resultado, más éxito tendrá la inteligencia empresarial. Además, el resultado no debe ser engañoso: si el usuario malinterpretara la información, tomaría malas decisiones, minando así el negocio de su empresa. BI prepara los datos listos para ser analizados, y los informes solo sirven para presentar esos datos de una manera que sea fácil de entender.

## Equipo del proyecto

La complejidad del sistema de BI requiere un equipo de personas con habilidades heterogéneas, tanto de TI como relacionadas con los procesos de negocio. Las figuras profesionales necesarias para la implementación de este sistema de BI se enumeran a continuación, donde junto a la figura profesional también se indica el rol desempeñado:

- **Ingeniero de datos** (Pau Papiol): se ocupará de las tareas de recolección, limpieza, transformación e integración de los datos. Trabaja con los datos al mínimo nivel de detalle, generalmente con tecnologías como Python, R, SQL o herramientas ETL.
- **Analista de datos** (Marc Faravelli): será el encargado del modelamiento matemático de los datos, disciplinas como análisis estadístico, técnicas de machine learning, análisis predictivo, clustering y reconocimiento de patrones. Domina técnicas para extracción de información desde las fuentes de datos.
- **Analista de negocio** (Jordi Poyato): se encarga de transmitir y hacer sentido de los datos al negocio, conoce y entiende la estrategia de la organización y su misión es guiar y apoyar con datos las decisiones que se deban tomar. Es importante el componente de visualización de datos en este rol y la capacidad de empatía con las personas del negocio.

- **Ingeniero de sistemas / experto en redes** (Diego Ceccatto): se encargará del mantenimiento de las redes informáticas globales y locales. Los sistemas más complejos con una topología de varios niveles a veces pueden fallar o dejar de funcionar. Su tarea será de asegurar el óptimo funcionamiento de cada computadora, así como instalar sistemas operativos, mantener varios componentes de los programas en condiciones adecuadas y reparar piezas y componentes que han fallado.

Cabe señalar que una misma persona puede desempeñar diferentes roles, siempre que tenga las habilidades. También enfatizar la importancia del experto empresarial y la retroalimentación de los gerentes y usuarios que utilizan los servicios ofrecidos por BI. Para crear un sistema de BI eficiente y efectivo, es esencial que sea el negocio que oriente al departamento de TI durante las fases de desarrollo de un sistema de BI y no al revés.

## Metodología

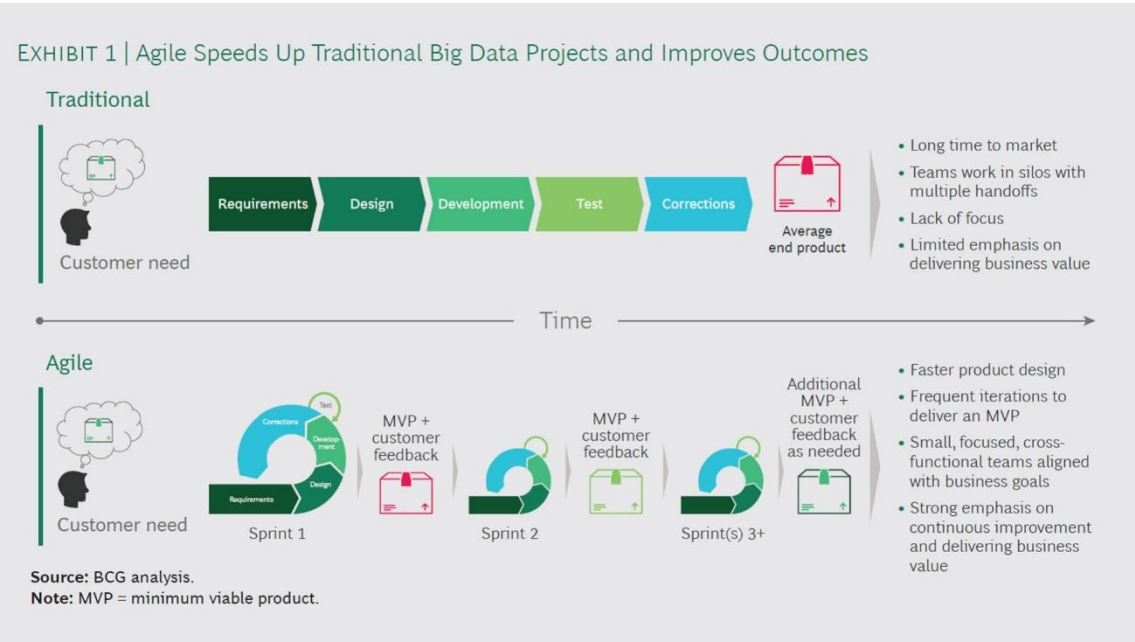
La BI tiene un coste general que puede ser muy alto. Si es oneroso implementar BI, es incluso más caro emprender un proyecto fallido, que generaría solo costes y no beneficios económicos. Por esta razón, es importante prestar mucha atención a cada fase del ciclo de vida del proyecto. El esquema de trabajo propuesto constituye un enfoque del diseño de BI que minimiza los riesgos de fracaso del proyecto. Este enfoque se basa en algunos puntos fijos:

- El análisis preliminar (valoración y justificación económica del proyecto).
- La actividad de planificación, con la que se determinan los tiempos y plazos de las fases posteriores.
- La construcción de prototipos.
- El proceso iterativo comienza con la recogida de requisitos e incluye, además de una fase analítica, el diseño del data warehouse y por último, la verificación del trabajo realizado.

Además, dentro del proceso iterativo, las fases para la implementación de OLAP, reporting y data mining se caracterizan por su propio ciclo de implementación.

Centrándonos en la metodología, la que se implementará será la Agile. Usar este método significa dividir el proyecto en partes fácilmente manejables y testeables individualmente, sin tener que esperar a que se complete la solución. Esto permite, por ejemplo, insertar correcciones durante la construcción como la introducción de fuentes de datos no previstas inicialmente, pero también abandonar el proyecto si aparecen mejores alternativas, ahorrando tiempo y dinero.

A continuación, podemos observar un cuadro de comparación entre la metodología tradicional y la agile:



Fuente: BCG

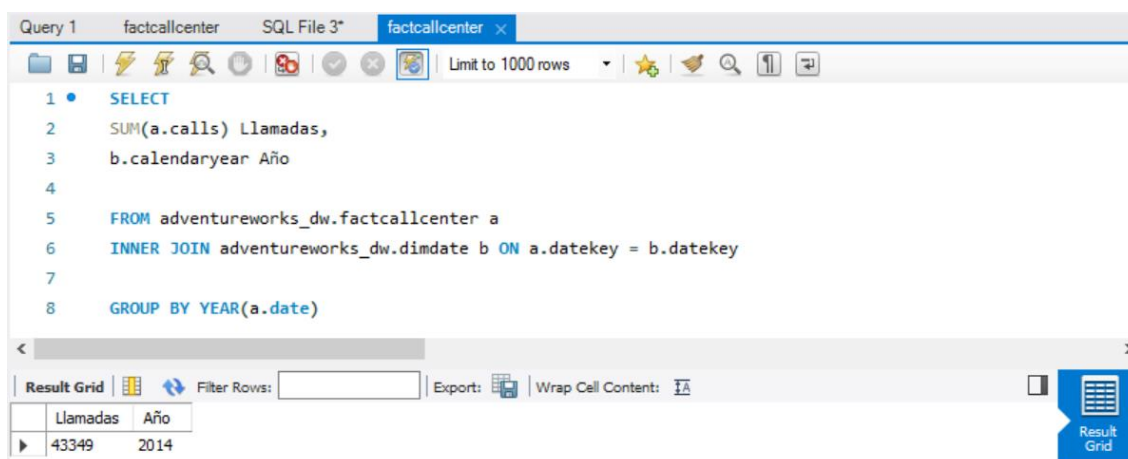
Otro aspecto para tener en consideración en cuanto a la efectividad de proyecto es que los equipos ágiles deben incluir diferentes talentos elegidos según sea necesario, como ya hemos visto en el apartado anterior.

## Fase 2 – Interpretación de datos con SQL

### Interpretando datos con lenguaje SQL

Departamento de Call Center (Tabla de hechos factcallcenter)

1. Escribe una consulta para obtener el total de llamadas (calls) agrupados por cada año. Ordena el resultado por el número de llamadas descendente.  
Nota: Deberás realizar un JOIN con otra tabla para extraer el año  
¿Qué año es en el que más llamadas se han realizado?



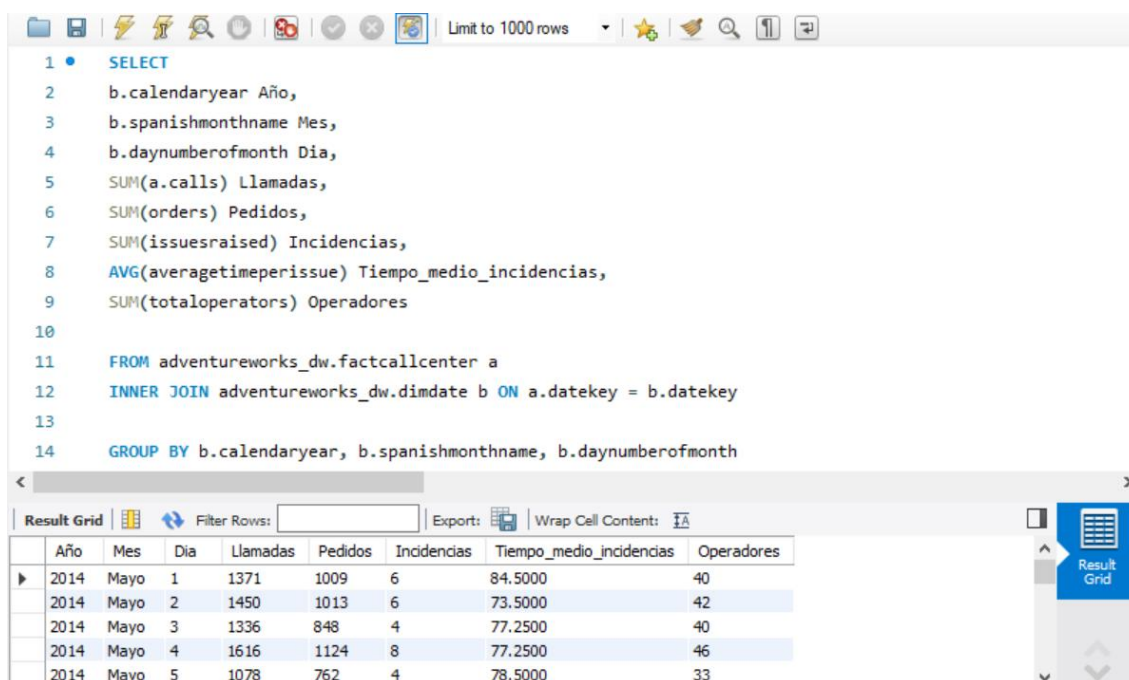
The screenshot shows a SQL query in the 'factcallcenter' window. The query is as follows:

```
1 • SELECT
2   SUM(a.calls) Llamadas,
3   b.calendaryear Año
4
5 FROM adventureworks_dw.factcallcenter a
6 INNER JOIN adventureworks_dw.dimdate b ON a.datekey = b.datekey
7
8 GROUP BY YEAR(a.date)
```

The result grid below the query shows the following data:

Llamadas	Año
43349	2014

2. Escribe una consulta obteniendo el número total de llamadas (calls), número de pedidos (orders), número de incidencias (issuesraised), tiempo medio de por incidencia (averagetimeperissue) y número de operadores (totaloperators) agrupados por Año, Mes y Día.



The screenshot shows a SQL query in the 'factcallcenter' window. The query is as follows:

```
1 • SELECT
2   b.calendaryear Año,
3   b.spanishmonthname Mes,
4   b.daynumberofmonth Día,
5   SUM(a.calls) Llamadas,
6   SUM(orders) Pedidos,
7   SUM(issuesraised) Incidencias,
8   AVG(averagetimeperissue) Tiempo_medio_incidencias,
9   SUM(totaloperators) Operadores
10
11 FROM adventureworks_dw.factcallcenter a
12 INNER JOIN adventureworks_dw.dimdate b ON a.datekey = b.datekey
13
14 GROUP BY b.calendaryear, b.spanishmonthname, b.daynumberofmonth
```

The result grid below the query shows the following data:

Año	Mes	Día	Llamadas	Pedidos	Incidencias	Tiempo_medio_incidencias	Operadores
2014	Mayo	1	1371	1009	6	84.5000	40
2014	Mayo	2	1450	1013	6	73.5000	42
2014	Mayo	3	1336	848	4	77.2500	40
2014	Mayo	4	1616	1124	8	77.2500	46
2014	Mayo	5	1078	762	4	78.5000	33



3. Escribe una consulta que obtenga el número de incidencias por número de semana y tipo de salario. Esta consulta debe obtener los datos para representar el siguiente gráfico.

```
1 • SELECT
2   SUM(issuesraised) Incidencias,
3   b.weeknumberofyear Semana,
4   a.wagetype Tipo_salario
5
6 FROM adventureworks_dw.factcallcenter a
7
8 INNER JOIN adventureworks_dw.dimdate b ON a.datekey = b.datekey
9
10 GROUP BY b.weeknumberofyear, a.wagetype
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Incidencias	Semana	Tipo_salario
▶	12	18	weekday
	4	18	holiday
	12	19	holiday
	28	19	weekday
	15	20	holiday
	27	20	weekday
	16	21	holiday
	19	21	weekday
	11	22	holiday
	18	22	weekday

Result Grid  
Form Editor  
Field Types

4. Turnos de trabajo: ¿Cuántos turnos de trabajo (shift) hay en el departamento? Escribe la consulta.

```
1 • SELECT
2   shift Turnos_trabajo
3 FROM adventureworks_dw.factcallcenter
4
5 GROUP BY shift
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Turnos_trabajo
▶	AM
	PM1
	PM2
	midnight

Result Grid  
Form

5. Escribe una consulta que obtenga el número de pedidos por turno de trabajo y número de semana. Esta consulta debe obtener los datos para representar el siguiente gráfico.

```

1 • SELECT
2   SUM(orders) Pedidos,
3   a.shift Turnos,
4   b.weeknumberofyear Semana
5
6 FROM adventureworks_dw.factcallcenter a
7
8 INNER JOIN adventureworks_dw.dimdate b ON a.datekey = b.datekey
9
10 GROUP BY a.shift, b.weeknumberofyear

```

	Pedidos	Turnos	Semana
▶	804	AM	18
	898	PM1	18
	740	PM2	18
	428	midnight	18
	1619	AM	19
	1813	PM1	19
	2404	PM2	19
	774	midnight	19
	1517	AM	20

#### Departamento de ventas por Internet (Tabla de hechos factinternetsales)

1. Escribe una consulta que obtenga el número de unidades y suma de las ventas agrupadas por año y mes. El resultado de la consulta debe satisfacer las necesidades para representar el siguiente gráfico.

```

1 • SELECT
2   b.calendaryear Año,
3   b.spanishmonthname Mes,
4   SUM(a.salesamount) VENTAS,
5   SUM(a.orderquantity) UNIDADES
6
7 FROM adventureworks_dw.factinternetsales a
8 INNER JOIN adventureworks_dw.dimdate b ON a.orderdatekey = b.datekey
9
10 GROUP BY b.calendaryear, b.spanishmonthname

```

	Año	Mes	VENTAS	UNIDADES
▶	2010	Diciembre	43421.03639999999	14
	2011	Enero	469823.91480000044	144
	2011	Febrero	466334.90300000004	144
	2011	Marzo	485198.65940000076	150
	2011	Abril	502073.8458000007	157
	2011	Mayo	561681.4758000008	174

2. Siguiendo el punto anterior, filtra el resultado de la consulta para los países “Canada” y “France”. Nota: Deberás hacer join con la tabla “dimesaleterritory”

```
1 • SELECT
2   b.calendaryear Año,
3   b.spanishmonthname Mes,
4   SUM(a.salesamount) VENTAS,
5   SUM(a.orderquantity) UNIDADES
6
7 FROM adventureworks_dw.factinternetsales a
8 INNER JOIN adventureworks_dw.dimdate b ON a.orderdatekey = b.datekey
9 INNER JOIN adventureworks_dw.dimesaleterritory c ON a.salesterritorykey = c.salesterritorykey
10
11 GROUP BY b.calendaryear, b.spanishmonthname, c.salesterritorycountry
12 HAVING salesterritorycountry = "France" or salesterritorycountry = "Canada"
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [fA](#)

Año	Mes	VENTAS	UNIDADES
2010	Diciembre	3578.27	1
2011	Enero	21965.4382	7
2011	Febrero	25340.428200000002	8
2011	Marzo	17891.35	5
2011	Abril	11929.7264	5
2011	Mayo	29846.0764	10

Result Grid  
Form Editor

## Selección de herramientas

Dentro de los dos tipos de herramientas principales en el BI: ETL y visualización, se han elegido respectivamente Knime y Power BI como los más adecuados para cumplir con las exigencias de este proyecto y de la empresa.

KNIME tiene diferentes ventajas:

- Es la solución de código abierto líder en innovación basada en datos que nos ayudará a descubrir el potencial oculto en nuestros datos, extraer nuevas ideas o predecir escenarios futuros. El aspecto de que sea código abierto es relevante en cuanto nos permite poder trabajar sin costes de mantenimiento y uso.
- Con más de 2000 módulos, cientos de ejemplos listos para usar, una gama completa de herramientas integradas y la más amplia selección de algoritmos avanzados disponibles, KNIME es la caja de herramientas perfecta para cualquier científico de datos.
- KNIME es rápido de instalar, fácil de escalar e intuitivo de usar. De hecho, su punto fuerte es que se trata de una plataforma gráfica para el análisis de datos basada en nodos y flujos de trabajo.

En cuanto a POWER BI, hay algunos aspectos destacables que nos decantan para su elección:

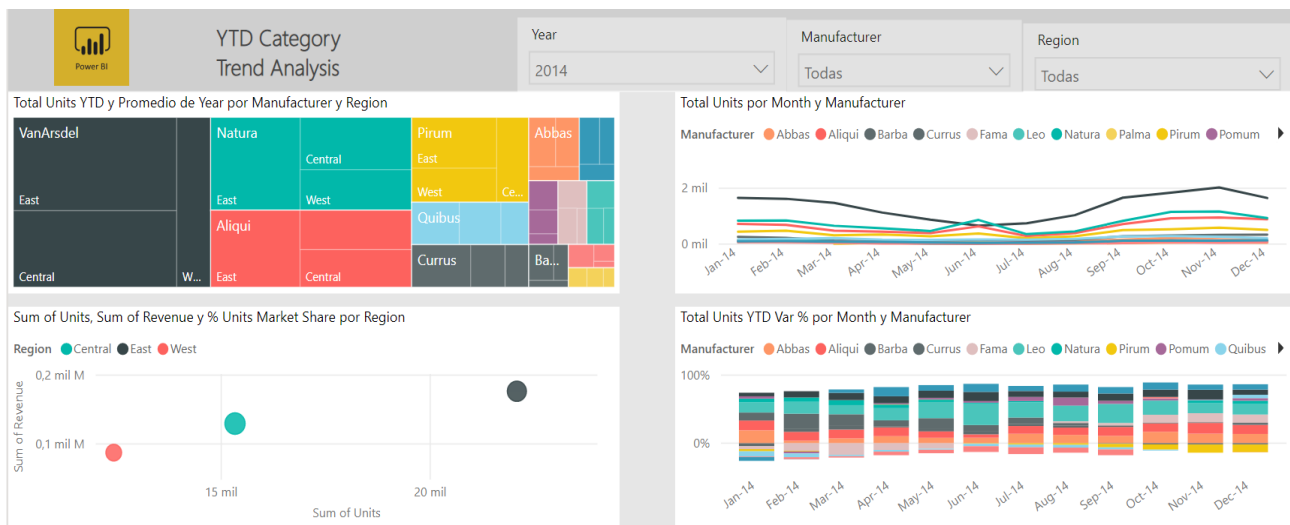
- Proporciona la plataforma perfecta para la BI ágil (nuestra metodología) porque nos permite configurar rápidamente nuevos enlaces de datos y crear nuevos informes analíticos. Esto es posible porque Power BI está estrechamente integrado con otros productos de Microsoft BI y con Excel, y cuenta con conectores integrados para fuentes de datos externas.
- Tiene soporte integrado para modelos tabulares, consultas de Power Query y orígenes de datos externos. En conjunto, estas oportunidades brindan acceso a todas las fuentes de datos que nuestros usuarios necesitan y abarcan datos relacionales, multidimensionales, estructurados o semiestructurados de fuentes locales, en la nube o web.
- Los informes y paneles de Power BI nos brindan una gran cantidad de oportunidades para explorar y visualizar datos, lo que nos permite usar los datos a nuestra manera. Esto nos permite individualizar nuestra experiencia, eligiendo la información de los datos que se ajuste a nuestros intereses y analizándola en el formato y dispositivo que preferimos.
- Las herramientas de Power BI reducen drásticamente el tiempo de aprendizaje de los nuevos usuarios, porque los informes y los paneles tienen una interfaz simple y consistente, que resultará familiar de inmediato para muchos usuarios de Excel de Power View o tablas dinámicas. Como alternativa, los usuarios avanzados que buscan herramientas de análisis más sofisticadas tienen la capacidad de crear consultas complejas y modelos de datos con Power BI Designer o Power Query y Power Pivot de Excel.

## Fase 3 – Aplicando los conocimientos en Power BI

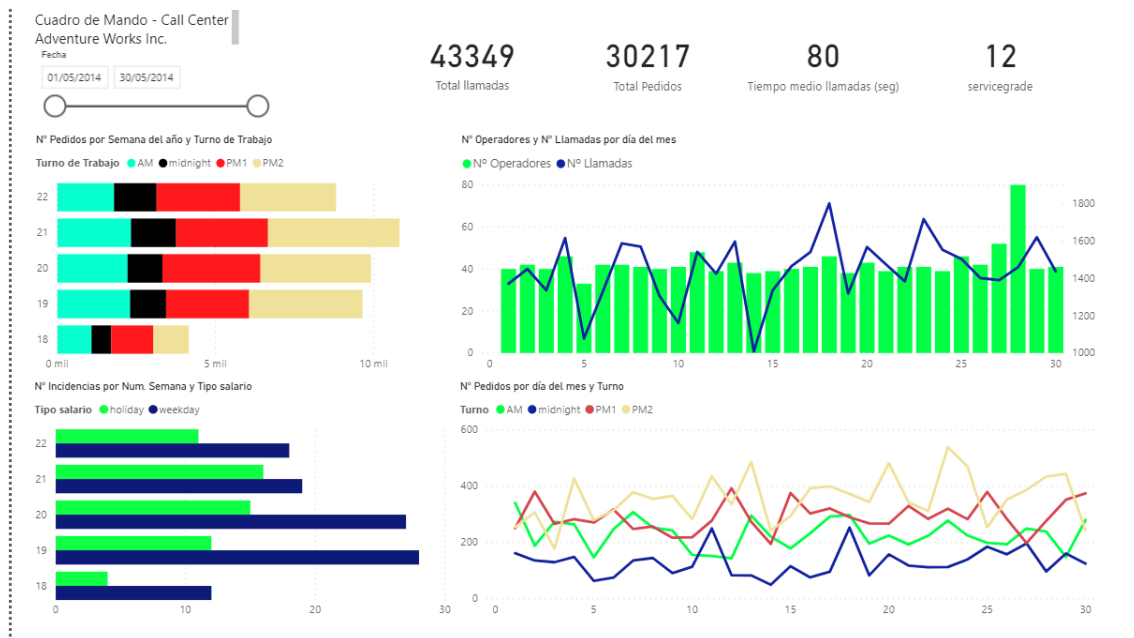
Partiendo de los conocimientos adquiridos en la clase 3 sobre Power BI, en esta fase del Caso Práctico se te pide que los pongas en práctica sobre un conjunto de datos concreto.

Como punto de partida tienes a tu disposición el fichero de Power BI “caso\_practico\_fase3.pbix”, el cual dispone de los conjuntos de datos necesarios para realizar esta fase del Caso Práctico.

### Cuadro de mandos



## Fase 4 – Implementación / Visualización de datos en Power BI en un caso real



### Parte opcional:

