

Explotación y Visualización

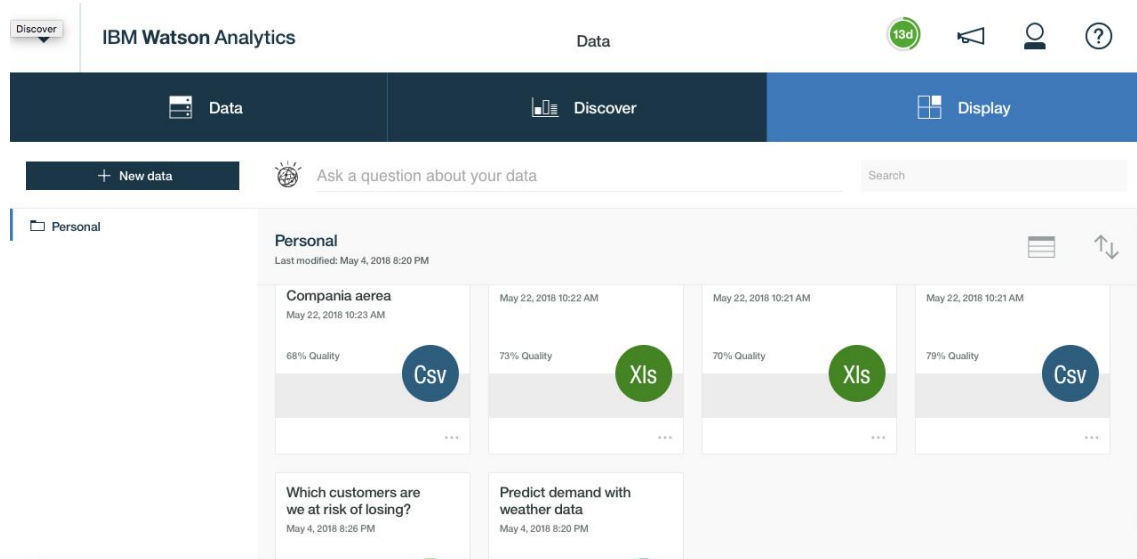
Práctica Watson Analytics

Caso de uso: Compañía aérea

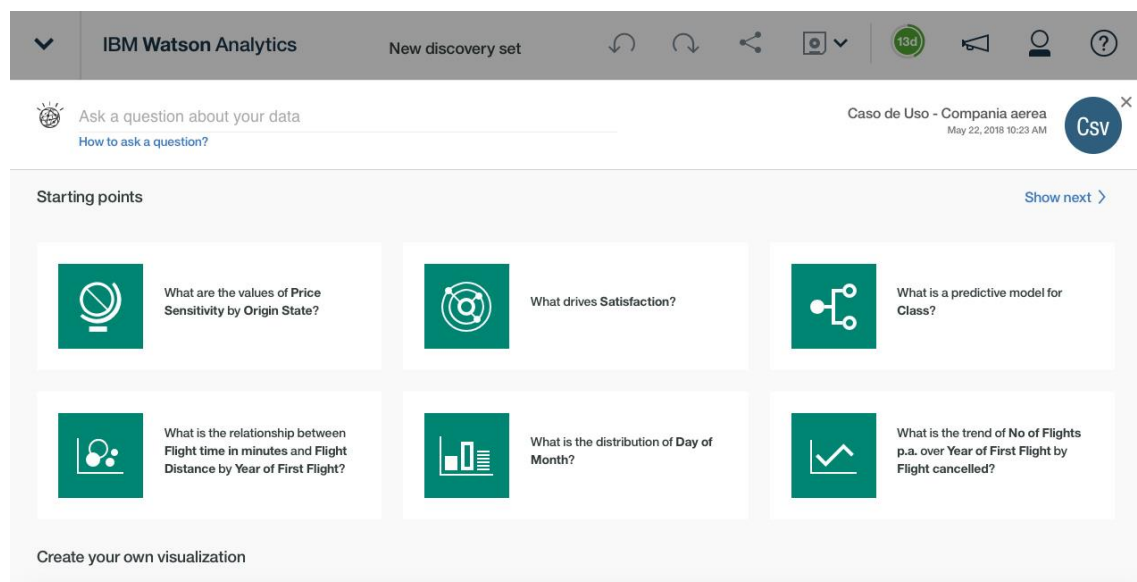
José Manuel Bustos Muñoz

1. Explorar que están diciendo los encuestados. ¿Cuál es la satisfacción de los viajeros con nuestra compañía, que clientes están no están contentos y que otras respuestas destacan es sus respuestas?

Lo primero sería cargar en la plataforma Watson Analytics el dataset con los datos a analizar.



Una vez se selecciona el dataset requerido, en este caso el de la compañía aérea, ya nos aparecen una sugerencia de posibles preguntas y análisis a realizar, en base a los atributos de los datos cargados.



Antes de empezar a realizar análisis en base a preguntas, también se puede en otra vista de los datos hacer un primer análisis exploratorio donde se muestran los atributos del dataset y los datos cargados.

Podemos ver como hay gran cantidad de atributos sobre los datos de viajes, parece que en el territorio de Estados Unidos. Viendo los datos que tenemos identificamos el atributo de “satisfacción” como la variable que querríamos evaluar para estudiar que variables influyen más en ella y como se podría llegar a predecir o mejorar sus valores aplicando principios de machine learning.

IBM Watson Analytics

*Caso de Uso - Compania aerea

13d

Caso de Us...

	Destination St...	Origin City	Origin State	Destination City	Year of First FL...	Day of Month	Flight date	Age	Age Range	Airline Code
Showing 1000 rows. Not all rows can be shown.										
	Alaska	Wrangell, AK	Alaska	Petersburg, AK	2004	10	03-10-2014	80	80+	AS
	Alaska	Petersburg, AK	Alaska	Wrangell, AK	2012	17	02-17-2014	46	40-49	AS
	Alaska	Wrangell, AK	Alaska	Petersburg, AK	2012	27	02-27-2014	73	70-79	AS
	Alaska	Petersburg, AK	Alaska	Wrangell, AK	2004	11	02-11-2014	23	20-29	AS
	Alaska	Wrangell, AK	Alaska	Petersburg, AK	2005	16	03-16-2014	78	70-79	AS
	Alaska	Wrangell, AK	Alaska	Petersburg, AK	2003	25	02-25-2014	38	30-39	AS
	Alaska	Wrangell, AK	Alaska	Petersburg, AK	2007	11	01-11-2014	37	30-39	AS
	Alaska	Petersburg, AK	Alaska	Wrangell, AK	2009	6	02-06-2014	41	40-49	AS
	Alaska	Petersburg, AK	Alaska	Wrangell, AK	2006	3	03-03-2014	43	40-49	AS
	Alaska	Wrangell, AK	Alaska	Petersburg, AK	2011	2	02-02-2014	53	50-59	AS
	Alaska	Petersburg, AK	Alaska	Wrangell, AK	2004	7	02-07-2014	69	60-69	AS
	Alaska	Petersburg, AK	Alaska	Wrangell, AK	2010	29	03-29-2014	49	40-49	AS
	Alaska	Wrangell, AK	Alaska	Petersburg, AK	2011	20	02-20-2014	42	40-49	AS
	California	Crescent City...	California	Arcata/Eurek...	2003	22	01-22-2014	43	40-49	OO
	California	Crescent City...	California	Arcata/Eurek...	2003	9	03-09-2014	32	30-39	OO
	California	Crescent City...	California	Arcata/Eurek...	2007	27	03-27-2014	69	60-69	OO
	California	Arcata/Eurek...	California	Crescent City...	2011	12	03-12-2014	45	40-49	OO
	California	Crescent City...	California	Arcata/Eurek...	2004	13	01-13-2014	32	30-39	OO
	California	Arcata/Eurek...	California	Crescent City...	2008	7	01-07-2014	51	50-59	OO

En esta pantalla se pueden ordenar los datos por alguna columna, realizar filtros o que se muestre un histograma de cada columna y algunos datos básicos de los datos cargados.

IBM Watson Analytics

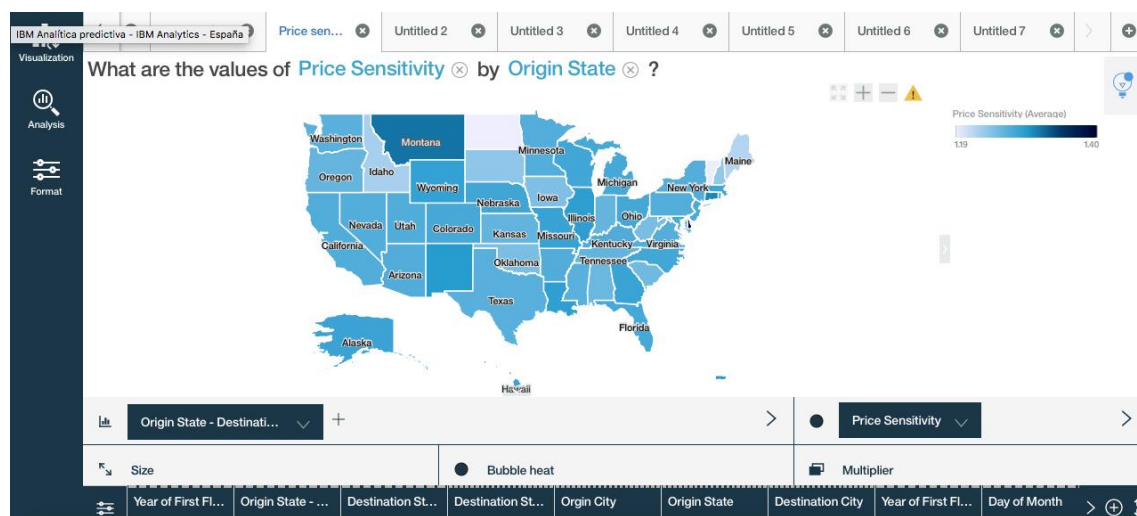
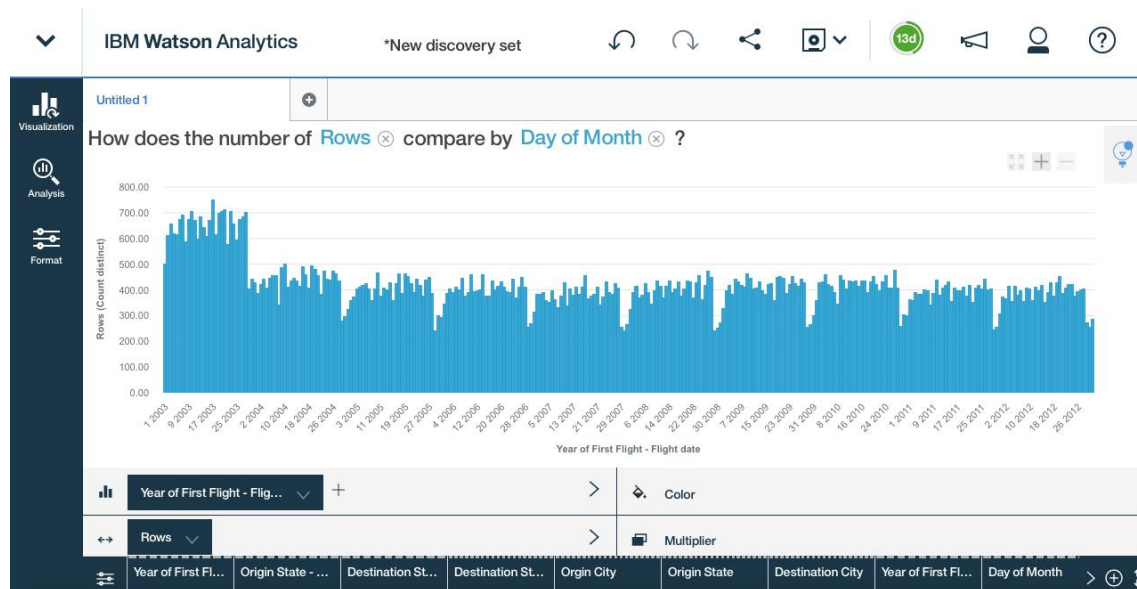
*Caso de Uso - Compania aerea

13d

Caso de Us...

	rival Delay I...	Departure Del...	Eating and Dri...	Flight time in ...	Shopping Am...	No of Flights p.a...	No. of other Lo...	% of Flight wit...	Price Sensitivity	Flight Distance
Showing 1000 rows. Not all rows can be shown.										
	49 Low Quality	49 Low Quality	55 Medium Quality	60 Medium Quality	51 Medium Quality	67 Medium Quality	63 Medium Quality	54 Medium Quality	68 Medium Quality	62 Medium Quality
	Missing Values: 2%	Missing Values: 2%	Missing Values: 0%	Missing Values: 2%	Missing Values: 0%	Missing Values: 0%	Missing Values: 0%	Missing Values: 0%	Missing Values: 0%	Missing Values: 0%
	0	0	30	18	60	17	3	12	\$1.00	84
	5	5	90	20	20	26	0	4	\$1.00	84
	0	0	50	20	90	39	0	7	\$1.00	84
	0	0	45	21	15	21	2	26	\$1.00	84
	0	2	122	25	82	33	2	12	\$1.00	84
	0	0	64	20	5	4	0	9	\$1.00	84
	16	17	70	26	0	6	1	16	\$1.00	84
	0	0	75	24	0	17	1	11	\$2.00	84
	0	6	75	20	30	21	2	2	\$2.00	84
	0	0	45	23	55	41	0	9	\$1.00	84
	0	0	35	23	120	0	0	9	\$1.00	84
	0	0	65	21	133	25	0	3	\$1.00	84
	0	0	55	20	0	7	0	2	\$1.00	84
	50	61	100	23	0	23	1	6	\$2.00	84

A la hora de realizar preguntas para hacer distintos análisis hay gran variedad de tipos de diagramas, por ejemplo podríamos realizar diagramas de barras o utilizar mapas para ver la información por regiones de un país.

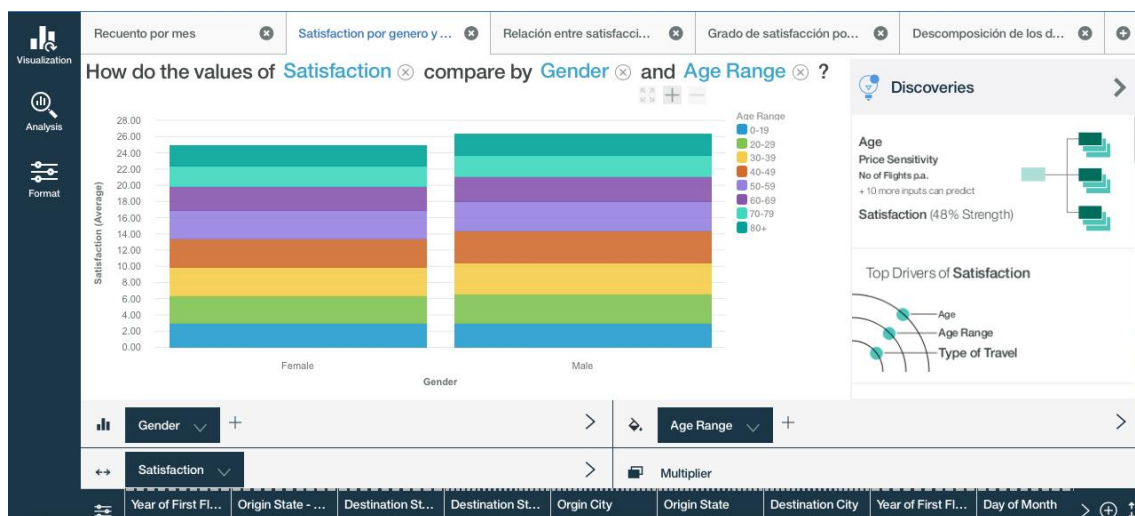


2. Identificar qué es lo que influye en la satisfacción para así poder desarrollar acciones para modificar la percepción de nuestros clientes.

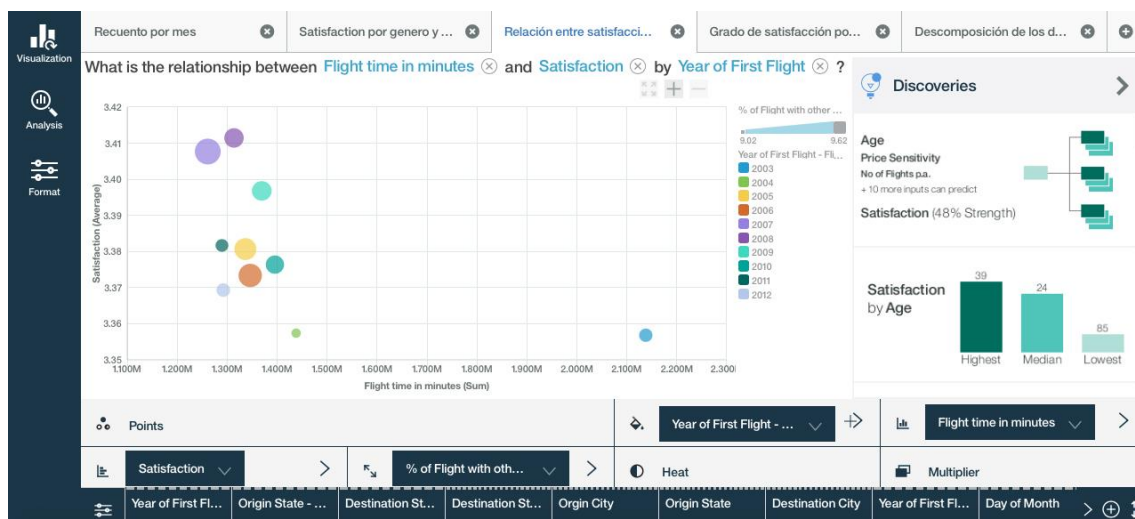
Como lo que queremos evaluar como variable target es el atributo “satisfacción” realizamos distintas preguntas con distintos tipos de diagramas para intentar analizar dicho atributo en base al resto de datos que tenemos y poder ver cuáles son las variables que más influyen en nuestra variable target.

Podemos realizar tanto preguntas propias como algunas que nos sugiera Watson. Además a medida que se realizan preguntas para hacer análisis también Watson puede ir indicándonos nuevas opciones a explorar.

Estudiamos los valores de satisfacción en base al género y al rango de edad. Por ejemplo vemos como no parece haber gran diferencia en base a estas variables.



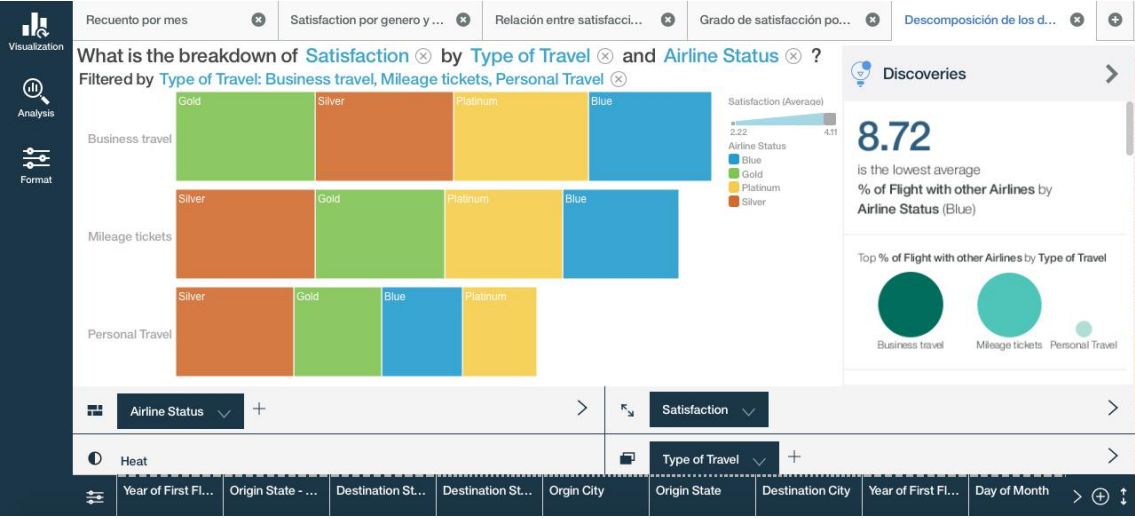
También estudiamos el grado de satisfacción por año y minutos de vuelo.



Parece interesante estudiar la satisfacción en base al año y a la clase del billete. Parece que hay mayor satisfacción en clase business, aunque no se aprecian unas grandes diferencias.



Otro análisis con otro tipo de diagrama es la descomposición en base al tipo de viaje y el status de la compañía, para también ver el peso del atributo de satisfacción.

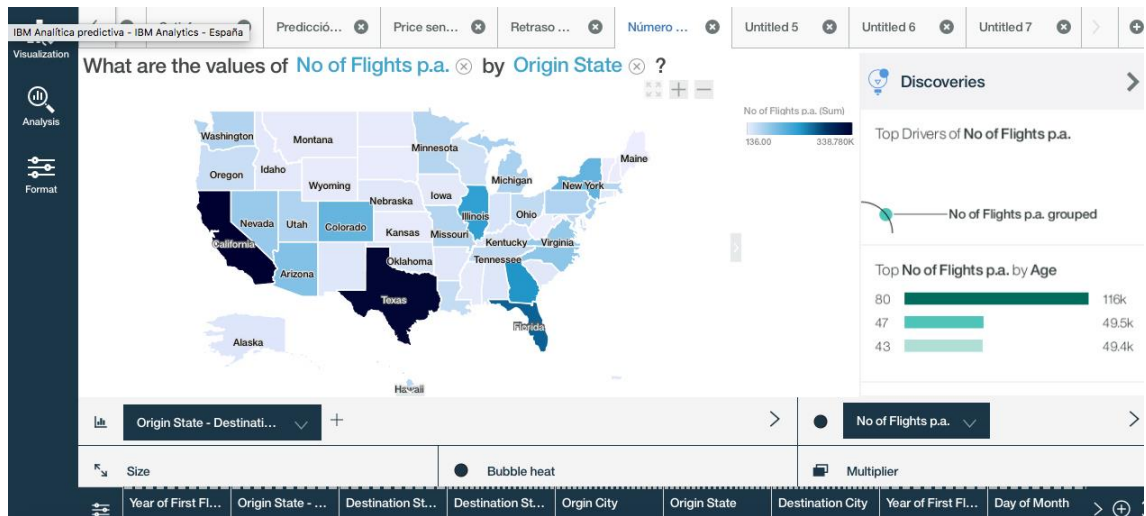


Evaluar con diagrama de barras la satisfacción a partir del rango de edad y el tipo de vuelo.



También se puede realizar otros análisis involucrando a otras variables no estudiadas para ver si aumenta la información estudiada y ayuda al modelo predictivo.

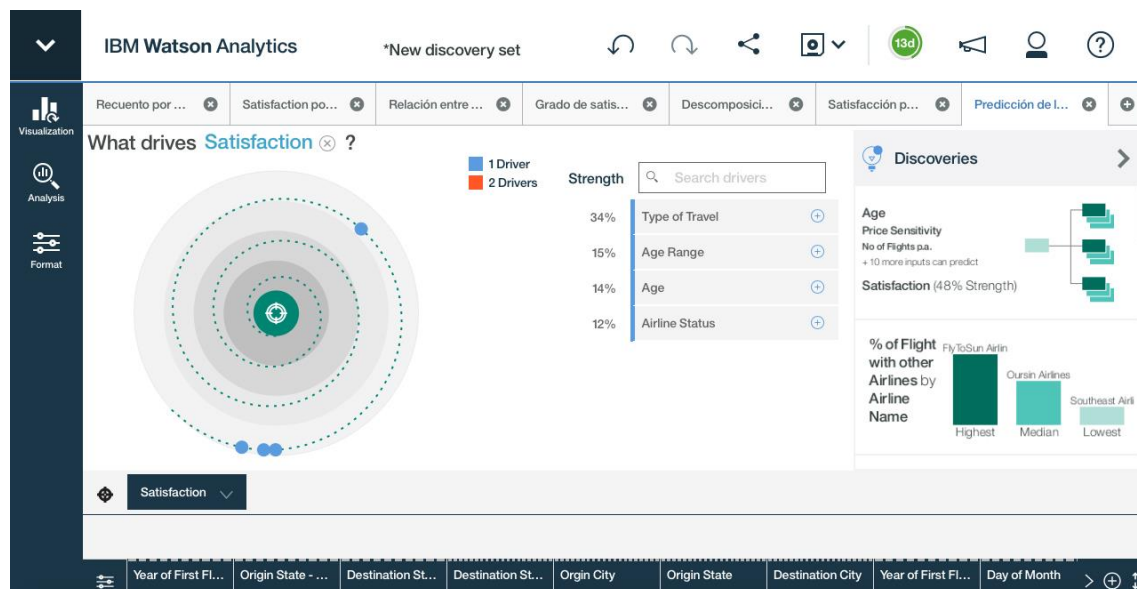
Por ejemplo el número de vuelos p.a por el estado origen, para ver posibles diferencias entre estados.



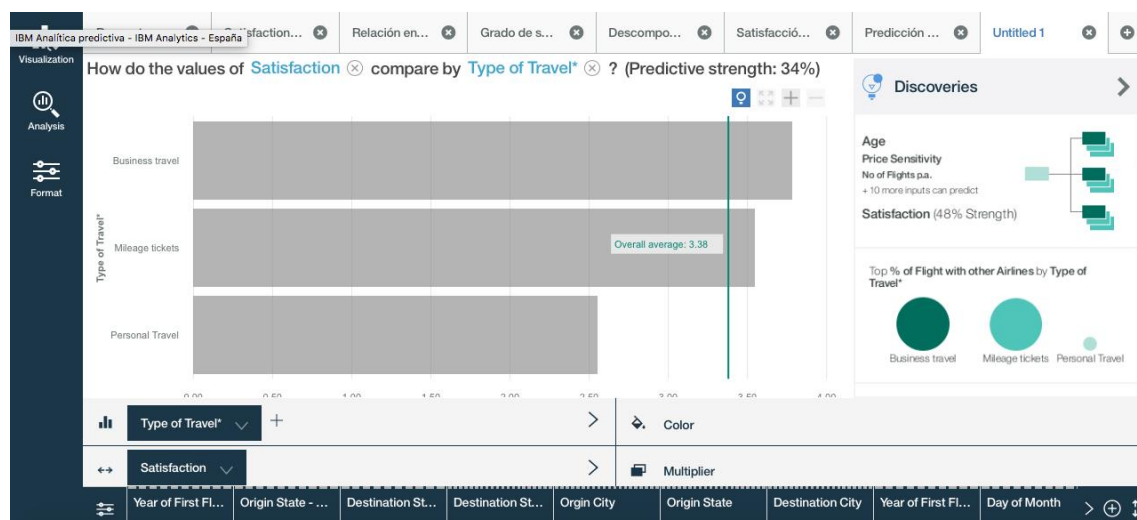
Al realizar las preguntas necesarias y que creamos suficientes, se puede estudiar con una espiral poniendo en el centro la variable target que variables son las que más influyen en nuestra variable a predecir.

En este caso nos indica que el tipo de viaje sería la variable predictora con mayor fuerza, y pone el porcentaje asociado. También aparecen las variables del rango de edad, la edad y el status de la aerolínea.

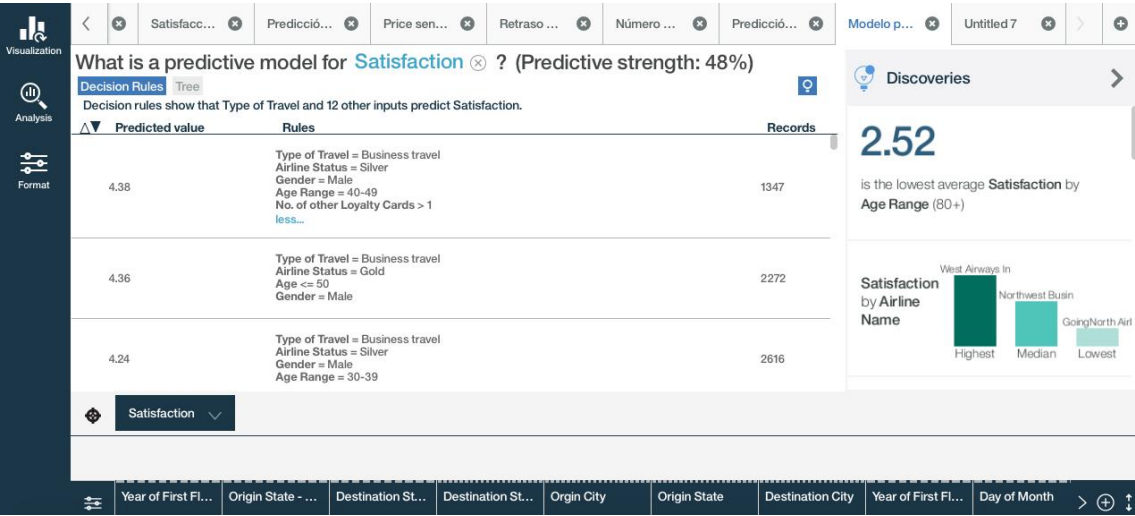
Además Watson también nos indica a la derecha posibles acciones a tomar.



Si pinchamos en alguna de las variables que ha indicado el modelo predictor aparece su información en relación a la variable target.

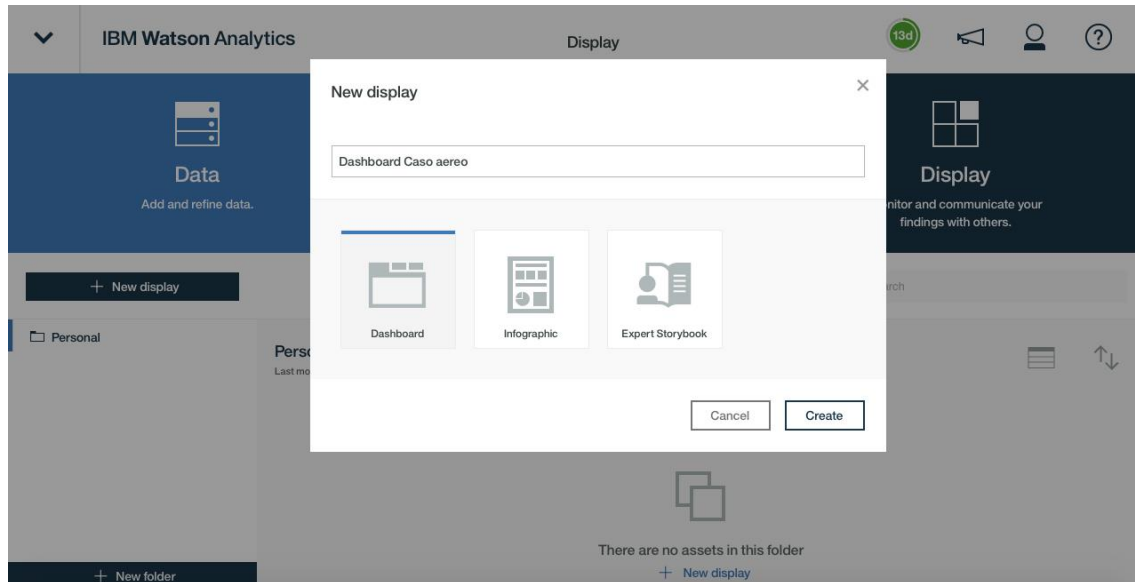


El modelo predictivo que nos aparece con una fuerza de un 48% nos indica reglas en base a valores de algunas de los atributos y el valor que llega a predecir para la satisfacción.

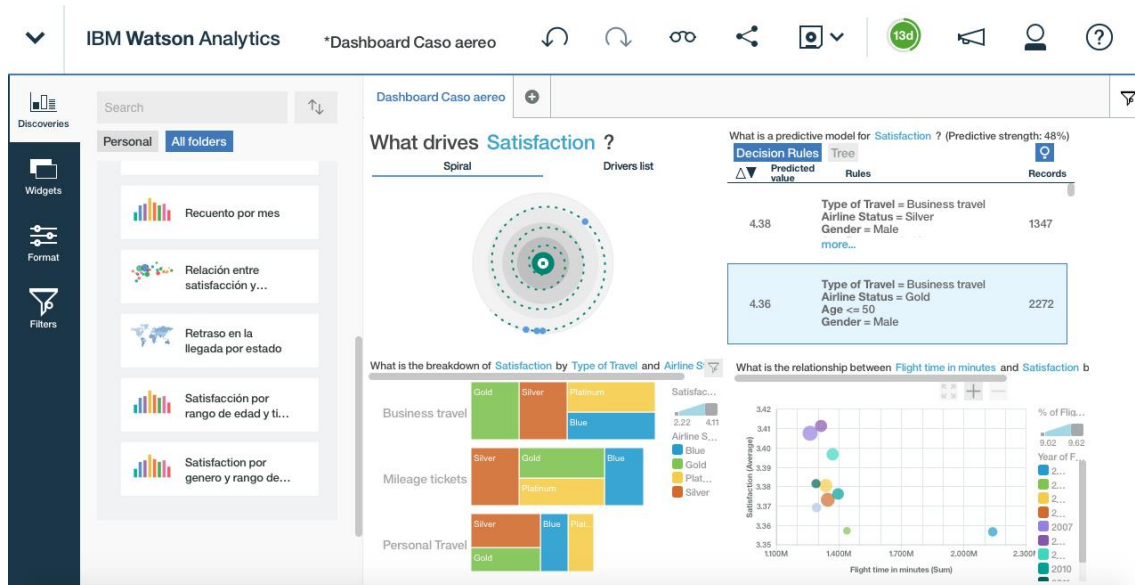


3. Desarrollar un cuadro de mando que muestre los resultados obtenidos, y utilizarlo para contar una historia de los que hemos descubierto, y de cualquier acción que vayamos a tomar.

Una de las opciones es realizar cuadros de mando donde se pueden colocar de muchas formas la información del análisis realizado y visualizar de forma rápida la información que queramos del trabajo realizado anteriormente.



Se crea un cuadro de mandos y por ejemplo colocamos la espiral para predecir el target, y los resultados del modelo predictor, además de dos de los diagramas realizados.



4. Analizar la opinión existente en los medios sociales de compañías aéreas más importantes de la actualidad española, e incluirlo en la presentación de los resultados.

Otra de las opciones de Watson es el análisis de sentimiento en las redes sociales.

Se accede a Watson Analytics for Social Media, y se cargan los datos. Entonces añadimos los topics y temas por los que queremos buscar en las redes sociales. En este caso al ser los datos de una compañía aérea ponemos topics y temas relacionados con el aeropuerto Barajas, la compañía aérea de Iberia, y conceptos como búsqueda de vuelos baratos o la comida en los aviones.

IBM Watson Analytics for Social Media

Caso aereo

Topics: 11 Topics

Themes: None

Dates: 04/22/2018 - 05/22/2018

Languages: 1 Selected

Sources: 7 Sources

Enter a topic

Search my topics

Iberia

Topic keywords: Iberia...

Aena

Topic keywords: Aena...

Barajas

Topic keywords: Baraj...

Aeropuerto Prat

Topic keywords: Aerop...

Espacio aereo

Topic keywords: Esao...

Vuelos baratos

Topic keywords: Vuelo...

Estimated documents 12K

Create data set

IBM Watson Analytics for Social Media

Caso aereo

Topics: Iberia selected

Themes: None

Dates: 04/22/2018 - 05/22/2018

Languages: 1 Selected

Sources: 7 Sources

Topics: Iberia

Topic keywords

eg. Pies

Iberia

Expand your search

Context keywords

eg. Bakery

Filter results by keyword

Exclude keywords

eg. Cutie

Remove unwanted content

Suggestions for Iberia

Refresh

Search

2017 song music songs 2016 movie live lyrics full remix dance videos 2018 trailer bitcoin episode love cover kids review

check ryanair seat passengers hours departure customers reserve policy seats pay reserved days online june airline window

Estimated documents 12K

Create data set

▼

IBM Watson Analytics for Social Media

Caso aereo

↓

13d

🔊

👤

?

📁 Topics

11 Topics

📁 Themes

6 Themes

📅 Dates

04/22/2018 - 05/22/2018

🌐 Languages

1 Selected

📁 Sources

7 Sources

Enter a theme

🔍 Search my themes

Comidas en aviones

Theme keywords: Comid...

Vuelos baratos

Theme keywords: Vuelo...

Turbulencias en avión

Theme keywords: Turbu...

Retraso en vuelos

Theme keywords: Retra...

Aeropuerto Adolfo Suarez barajas Madrid

Theme keywords: Aero...

Transporte del centro al aeropuerto

Theme keywords: Trans...

Estimated documents

12K

Create data set

Por último se puede seleccionar las fuentes en las que queremos realizar la búsqueda.

▼

IBM Watson Analytics for Social Media

Caso aereo

↓

13d

🔊

👤

?

📁 Topics

11 Topics

📁 Themes

8 Themes

📅 Dates

04/22/2018 - 05/22/2018

🌐 Languages

2 Selected

📁 Sources

7 Sources

Reddit

Reddit conversations and communities that feature over 270 million monthly users.

Enabled ☒

Forums

Content from more than 5 million message boards and forums.

Enabled ☒

Reviews

Comments and reviews from hundreds of review sites.

Enabled ☒

Facebook pages

View pages

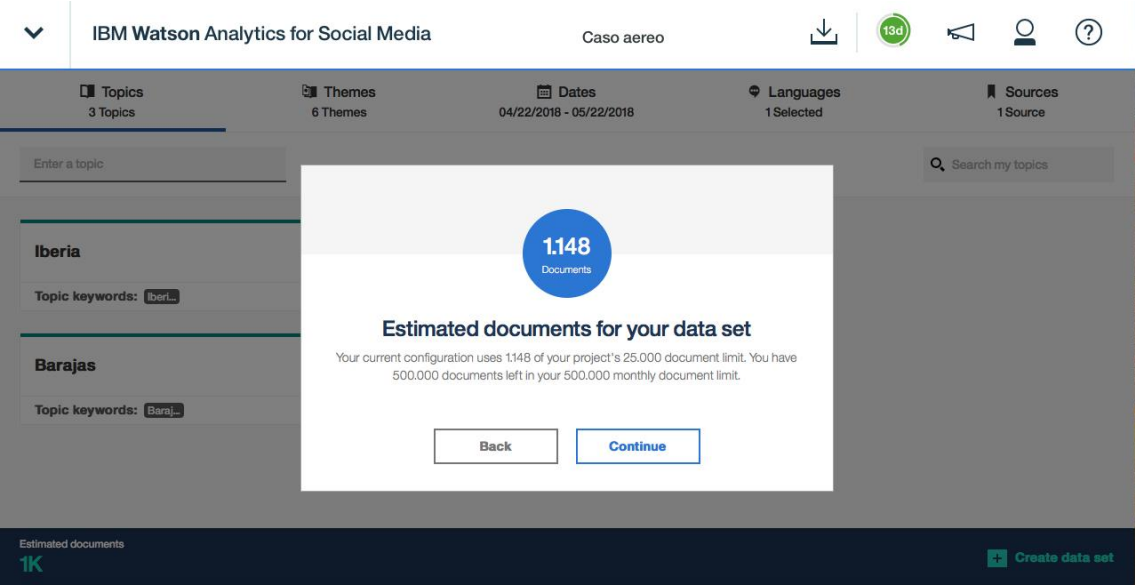
Enabled ☒

Estimated documents

31K

Create data set

Para probar seleccionamos sólo dos topics y algunos temas para que no sea muy grande el tamaño de documentos encontrados.



Una vez terminado el análisis podemos visualizar los resultados, y por ejemplo en la imagen ver como se han encontrado en Facebook comentarios relacionados con Iberia. Y podría estudiarse los comentarios positivos y negativos hacia la compañía aérea en cuestión.

