

Propuesta proyecto Shiny

Herramienta de análisis post-modelación

Jaime Andrés Fierro Aponte

David Alejandro Ibarra Parada

Laura Alejandra Melo Buitrago

Laura Daniela Rodríguez Molina

Contenido

1. Introducción al tema
2. Objetivos
3. Descripción de los datos
4. Propuesta de metodología
5. Resultados esperados

Introducción al tema

La empresa

GroupM es una multinacional ofrece servicios de marketing a clientes a nivel global.

The logo for GroupM, featuring the word "group" in a dark blue, lowercase, sans-serif font, followed by a stylized "m" in a lighter blue, lowercase, sans-serif font.

Al ser una empresa grande, se divide en agencias y subagencias que se enfocan en temas específicos.



M Services (ahora Nexus), se encarga de hacer modelos econométricos para determinar cuáles canales de medios explican las ventas.

The logo for Nexus, featuring the word "nexus" in a dark blue, lowercase, sans-serif font.

El equipo

El equipo Data Analytics de Nueva York con filial en Colombia, se encarga de:

- Procesar los datos que envían los clientes.
- Hacer modelos que permitan establecer qué tanto aporta cada canal de medios al KPI de interés del cliente.
- Generar recomendaciones de inversión para los clientes para mejorar su retorno a la inversión en publicidad de medios.
- Hacer modelos que permitan determinar en el nivel más granular posible, cuáles estrategias de publicidad son más rentables.

El problema

Después de hacer el modelo, obtenemos un conjunto de archivos con distintas partes de los resultados. Para evaluar el modelo, calcular las contribuciones, los retornos a la inversión y demás resultados de interés, se suele hacer con el procesamiento de dichos datos en Excel. Lo que toma mucho tiempo.



Objetivos

Objetivo general

















Generar una herramienta que presente los resultados que típicamente se obtienen mediante procesos repetitivos en Excel o en archivos R. Esto para optimizar el uso del tiempo y minimizar los errores humanos que se podrían cometer al hacer el procesamiento “manualmente”. Esto es, automatizar los procesos repetitivos.

Objetivos específicos

- Automatizar la evaluación del modelo realizado a partir del coeficiente de determinación calculado a través de individuos.
- Automatizar la visualización de los valores observados vs los valores estimados por el modelo.
- Automatizar el cálculo de los residuales por individuo y su visualización.
- Automatizar el cálculo de las contribuciones por grupos de variables según la necesidad del usuario.
- Automatizar el cálculo de los retornos a la inversión a partir de las contribuciones previamente calculadas.

Descripción de los datos

Archivo ZIP

↑  Results for Deep Dives 2.zip - archivo ZIP, tamaño descomprimido 37.825.130 bytes					
Nombre	Tamaño	Comprimido	Tipo	Modificado	CRC32
 ..			Carpeta de archivos		
 Model Decomposition.HTML	5.833.255	629.529	Microsoft Edge HT...	6/09/2022 5:15 ...	F8BA347F
 Model Contributions.csv	27.457.527	2.234.685	Archivo de valores s...	6/09/2022 5:15 ...	A145E22E
 Residual-Seasonality.HTML	580.897	76.855	Microsoft Edge HT...	6/09/2022 5:15 ...	A0D6A14B
 Model Statistics.HTML	6.652	2.043	Microsoft Edge HT...	6/09/2022 5:15 ...	93478953
 Model AVM.csv	696.508	174.455	Archivo de valores s...	6/09/2022 5:15 ...	89CE336B
 Variables Statistics.csv	311.151	12.341	Archivo de valores s...	6/09/2022 5:15 ...	7F4075E1
 Variables Statistics Outside.HTML	3.644	1.484	Microsoft Edge HT...	6/09/2022 5:15 ...	7898BCD9
 Model Statistics Outside.HTML	3.640	1.485	Microsoft Edge HT...	6/09/2022 5:15 ...	57E29391
 Model Contributions.HTML	1.462.842	319.026	Microsoft Edge HT...	6/09/2022 5:15 ...	479174CD
 Variables Statistics Outside.csv	30	30	Archivo de valores s...	6/09/2022 5:15 ...	435BA9B8
 Model Statistics Outside.csv	30	30	Archivo de valores s...	6/09/2022 5:15 ...	435BA9B8
 Variables Statistics.HTML	837.474	21.435	Microsoft Edge HT...	6/09/2022 5:15 ...	249CDE4B
 Actual vs Model.HTML	630.978	132.324	Microsoft Edge HT...	6/09/2022 5:15 ...	11C52F33
 Model Statistics.csv	502	312	Archivo de valores s...	6/09/2022 5:15 ...	0AF59235

Model AVM en Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Iteration	Geography	Product	Campaign	Outlet	Creative	Period	DepVar	DepVarAvg	Model	Forecast	Residual
2	1	YETI website - PAC					5/12/2016		104.019.296.207.974			
3	1	YETI website - PAC					28/08/2017		104.019.296.207.974			
4	1	YETI website - PAC					4/09/2017		104.019.296.207.974			
5	1	YETI website - PAC					2/10/2017		104.019.296.207.974			
6	1	YETI website - PAC					16/10/2017		104.019.296.207.974			
7	1	YETI website - PAC					30/10/2017		104.019.296.207.974			
8	1	YETI website - PAC					6/11/2017		104.019.296.207.974			
9	1	YETI website - PAC					13/11/2017		104.019.296.207.974			
10	1	YETI website - PAC					20/11/2017		104.019.296.207.974			
11	1	YETI website - PAC					27/11/2017		104.019.296.207.974			
12	1	YETI website - PAC					4/12/2017		104.019.296.207.974			
13	1	YETI website - PAC					11/12/2017		104.019.296.207.974			
14	1	YETI website - PAC					18/12/2017		104.019.296.207.974			
15	1	YETI website - PAC					25/12/2017		104.019.296.207.974			
16	1	YETI website - PAC					1/01/2018	835.418.473.035.273	104.019.296.207.974	743.668.890.004.332		0.917495830309407
17	1	YETI website - PAC					8/01/2018	818.072.415.478.991	104.019.296.207.974	790.639.704.445.478		0.27432711033513
18	1	YETI website - PAC					15/01/2018	872.701.155.545.034	104.019.296.207.974	850.788.433.407.283		0.21912722137751
19	1	YETI website - PAC					22/01/2018	808.632.411.776.617	104.019.296.207.974	866.011.354.444.384		-0.573789426677665
20	1	YETI website - PAC					29/01/2018	824.339.035.521.429	104.019.296.207.974	928.810.684.811.116		-104.471.649.289.687

Variables Model AVM



Geography (Categorica): Dimensión transversal de los datos.



Period (Fecha): Fecha en que se realiza la venta.



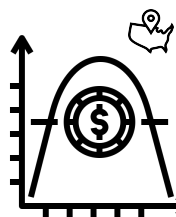
DepVar (Continua): Información real de la variable dependiente.



Model (Continua): Estimación de la variable dependiente.



Residual (Continua): Diferencia entre el valor de la estimación y el valor real.



DepVarAvg (Continua): Promedio de la información real de la variable dependiente y zona geográfica de venta.

Model Contributions en Excel

	A	B	C	D	E	F	G	H	I	J	K
1	Iteration	Geography	Product	Campaign	Outlet	Creative	VariableName	Period	VarCon	VarConROI	VariableValue
2	1	YETI website - PAC					Intercept	1/01/2018	6.642638861	-2.125202348	1
3	1	YETI website - PAC					Intercept	8/01/2018	6.642638861	-2.125202348	1
4	1	YETI website - PAC					Intercept	15/01/2018	6.642638861	-2.125202348	1
5	1	YETI website - PAC					Intercept	22/01/2018	6.642638861	-2.125202348	1
6	1	YETI website - PAC					Intercept	29/01/2018	6.642638861	-2.125202348	1
7	1	YETI website - PAC					Intercept	5/02/2018	6.642638861	-2.125202348	1
8	1	YETI website - PAC					Intercept	12/02/2018	6.642638861	-2.125202348	1
9	1	YETI website - PAC					Intercept	19/02/2018	6.642638861	-2.125202348	1
10	1	YETI website - PAC					Intercept	26/02/2018	6.642638861	-2.125202348	1
11	1	YETI website - PAC					Intercept	5/03/2018	6.642638861	-2.125202348	1
12	1	YETI website - PAC					Intercept	12/03/2018	6.642638861	-2.125202348	1
13	1	YETI website - PAC					Intercept	19/03/2018	6.642638861	-2.125202348	1
14	1	YETI website - PAC					Intercept	26/03/2018	6.642638861	-2.125202348	1
15	1	YETI website - PAC					Intercept	2/04/2018	6.642638861	-2.125202348	1
16	1	YETI website - PAC					Intercept	9/04/2018	6.642638861	-2.125202348	1
17	1	YETI website - PAC					Intercept	16/04/2018	6.642638861	-2.125202348	1
18	1	YETI website - PAC					Intercept	23/04/2018	6.642638861	-2.125202348	1
19	1	YETI website - PAC					Intercept	30/04/2018	6.642638861	-2.125202348	1
20	1	YETI website - PAC					Intercept	7/05/2018	6.642638861	-2.125202348	1
21	1	YETI website - PAC					Intercept	14/05/2018	6.642638861	-2.125202348	1
22	1	YETI website - PAC					Intercept	21/05/2018	6.642638861	-2.125202348	1
23	1	YETI website - PAC					Intercept	28/05/2018	6.642638861	-2.125202348	1
24	1	YETI website - PAC					Intercept	4/06/2018	6.642638861	-2.125202348	1
25	1	YETI website - PAC					Intercept	11/06/2018	6.642638861	-2.125202348	1
26	1	YETI website - PAC					Intercept	18/06/2018	6.642638861	-2.125202348	1
27	1	YETI website - PAC					Intercept	25/06/2018	6.642638861	-2.125202348	1
28	1	YETI website - PAC					Intercept	2/07/2018	6.642638861	-2.125202348	1
29	1	YETI website - PAC					Intercept	9/07/2018	6.642638861	-2.125202348	1

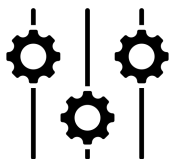
Variables Model Contributions



Geography (Categórica):
Dimensión transversal de los
datos.



Period (Fecha): Fecha en que se
realiza la venta.



VariableName (Categórica):
Información real de la venta



Product (Categórica): División del
producto.



Campaign (Categórica): Tipo de
campaña publicitaria.



Outlet (Categórica): Canal de
distribución de publicidad.



Creative (Categórica): Tipo de
anuncio de publicidad.

Variables Model Contributions



VarCon (Continua): Contribuciones calculadas para cada variable.



VarConROI (Continua): Retornos de la inversión calculados a partir de las contribuciones.



VariableValue (Continua): Valor de la variable.

Propuesta metodológica

Model AVM



Para alcanzar los objetivos propuestos, se establece una ruta que contiene la siguiente metodología:

- Se parte del archivo .zip que se descarga de la plataforma Analytics Workbench (AWB).
- Se selecciona el archivo llamado "Model AVM.csv".
- Se toman las variables *DepVar*, *Model* y *Residual* y se aplica exponente a cada una.
- Se calcula el coeficiente de determinación R^2 .
- Se grafica *DepVar* y *Model* para poder comprarlas.
- Se grafican los residuales por la variable "Geography"



Model Contributions



Para alcanzar los objetivos propuestos, se establece una ruta que contiene la siguiente metodología:

- Se parte del archivo .zip que se descarga de la plataforma Analytics Workbench (AWB).
- Se selecciona el archivo llamado "Model Contributions.csv".
- Se agrupan las categorías de la variable *VariableName* según se requiera.
- Se transforman los valores de la variable *VarCon*.
- Se calculan los ROIs a partir de las contribuciones transformadas.

Herramientas de R a utilizar

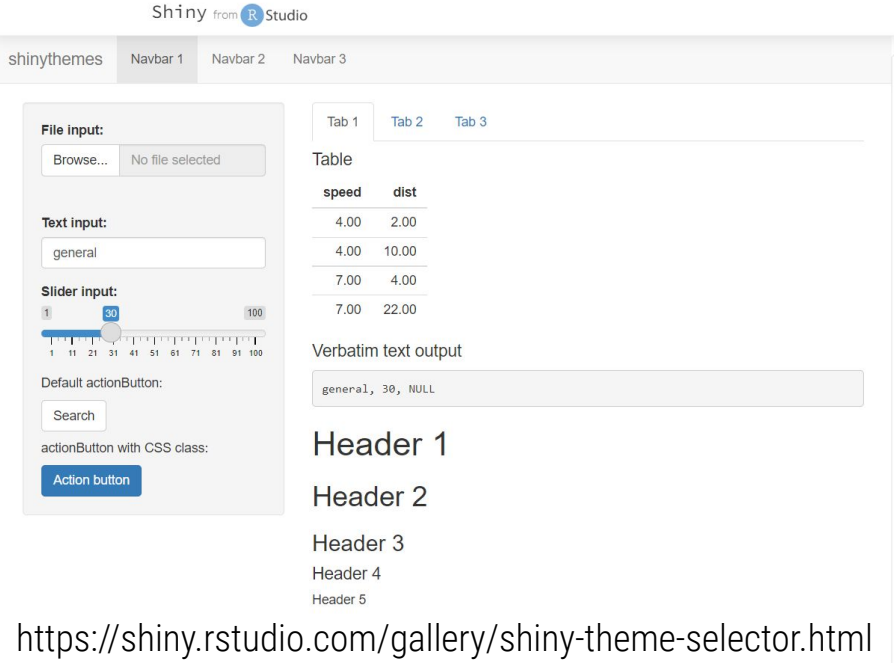
- Shiny: Para construir la interfaz.
- dplyr: Para manipular los datos (seleccionar las variables, hacer los filtros, las agrupaciones, renombrar variable).
- tidyr: Para ordenar los datos, manejar datos nulos, transformar la estructura de datos.
- ggplot2: Para hacer las gráficas.
- Plotly: Para gráficos interactivos.



Resultado esperado

Resultado esperado

Un shiny que permita cargar el archivo .zip, que calcule y grafique el desempeño del modelo y los residuales por geografía. Además, que transforme y agregue las contribuciones por deseados por el usuario y a partir de allí calcule los ROIs.



<https://shiny.rstudio.com/gallery/shiny-theme-selector.html>

	Total contribution	Contribution %	Spend	Latest 52 week ROI
Ambassador Impressions	\$ 7,186,400.51	13.89%	\$ 2,129,694.00	\$ 3.37
Digital Impressions	\$ 4,495,784.70	8.69%	\$ 3,679,489.00	\$ 2.11
Email Send Volume	\$ 2,097,007.16	4.05%	\$ 915,471.00	\$ 0.98
Print Impressions	\$ 1,120,092.69	2.16%	\$ 428,411.00	\$ 0.53
Social Impressions	\$ 255,829.74	0.49%	\$ 3,574,477.00	\$ 0.12
Sponsorship Impressions	\$ 141,219.63	0.27%	\$ 2,585,349.00	\$ 0.07

