



DATABRIDGE

SMART SOLUTIONS

CONTENIDO

KPIS

DASHBOARD INTERACTIVO

MODELO ML: SENTIMIENTO

MODELO ML: CLUSTERING



cronograma restante:

SEMANA 3	Conexion BigQuery con PowerBi					
	Conexion BigQuery con python					
	Dashboards					
	KPI's					
	Modelado ML					
	Analisis de sentimiento					
	Documentacion (feature engineer)					
	Documentacion (analisis)					
SEMANA 4	Documentacion a GitHub					
	Storytelling					
	Presentaciones (slides, videos)					
	Analisis de funcionalidad					
	Test de ML funcionando					

KPI	DESCRIPCION	FORMULA	FRECUENCIA
Volumen de reviews (popularidad)	aumentar en un 5% la cantidad de reviews respecto al mes anterior	$((\text{cant. reviews del mes actual} - \text{cant. reviews del mes anterior}) / \text{cant. reviews del mes anterior})$	Mensual
Proporcion positiva de reviews (baja negatividad)	La de reviews positivas debe mantenerse en 5:1 sobre las negativas. Considerando Review	$\text{cant. reviews positivas} / \text{cant. reviews negativas}$	Mensual
Satisfaccion del cliente (mejora del servi-	Aumentar a 3% el promedio de calificaciones	$((\text{promedio calificacion actual} - \text{promedio calificacion anterior}) / \text{promedio calificacion})$	Mensual
Aumento de catidad de locales por estado (aumento de	Aumentar en un 2% la cantidad de negocios por Estado	$((\text{cantidad negocios actual} - \text{cantidad de negocios anterior}) / \text{cantidad de negocios})$	Mensual
Rubro belleza vs rubro restaurant	Comparar el promedio de calificaciones del rubro belleza contra el rubro restaurant	$((\text{Promedio Calificaciones Belleza} - \text{Promedio Calificaciones Restaurant}) / \text{Promedio Calificaciones})$	Mensual

DASHBOARD

Lorem ipsum



MACHINE LEARNING

ANÁLISIS DE SENTIMIENTOS

Aplicación de técnicas de **procesamiento de lenguaje natural (NLP)** para analizar el sentimiento de las reseñas y clasificarlas en positivas, negativas o neutrales. Haciendo uso de la librería `SentimentIntensityAnalyzer` del conjunto `nlk.sentiment`. La cual genera una nueva columna donde se clasifica cada reseña, reemplazando/traduciendo así la reseña misma a su categoría representante.

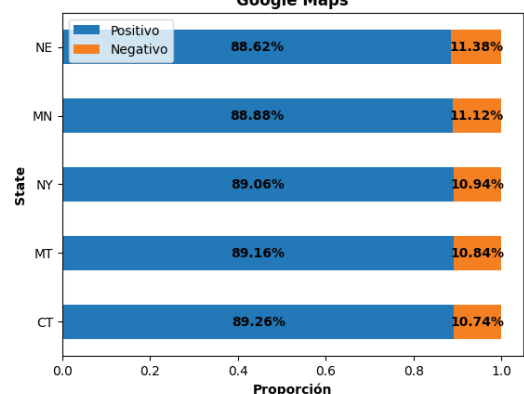
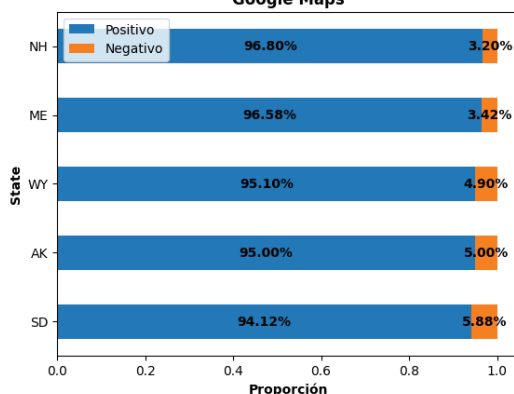
Es así como podemos ordenar y filtrar para dar a conocer cuales Estados se encuentran mas contentos con el servicio y cuales no.

Google Maps - RESULTADO TABLA:

	Nombre_Local	Hora	Sentimiento	Fecha	Servicio	Estado
0	SmartStyle Hair Salon	14:02:59	Positivo	2019-03-09	Hair salon	CO
1	Cactus Curlers Beauty Salon	23:32:13	Positivo	2020-03-18	Beauty salon	AZ
2	Cactus Curlers Beauty Salon	15:23:49	Positivo	2020-12-04	Beauty salon	AZ
3	Cactus Curlers Beauty Salon	13:17:33	Positivo	2019-06-06	Beauty salon	AZ
4	Sophie Nails	15:21:36	Positivo	2019-08-05	Nail salon	NC
...
64841	Nails By Anh	17:07:59	Positivo	2020-09-23	Nail salon	FL
64842	Stacy Taylor Salon	14:46:36	Positivo	2021-04-16	Hair salon	FL
64843	Stacy Taylor Salon	12:50:27	Positivo	2018-01-06	Hair salon	FL
64844	Stacy Taylor Salon	08:07:39	Positivo	2020-10-03	Hair salon	FL
64845	Patches & Stitches	11:29:00	Positivo	2019-12-29	Quilt shop	AL

Esta tabla de hechos representa las características y resultados de cada una de las reviews, filtradas del rubro belleza y estética, conjunto a una nueva columna llamada Sentimiento donde se expresa si la review fue Positiva, Neutral o Negativa

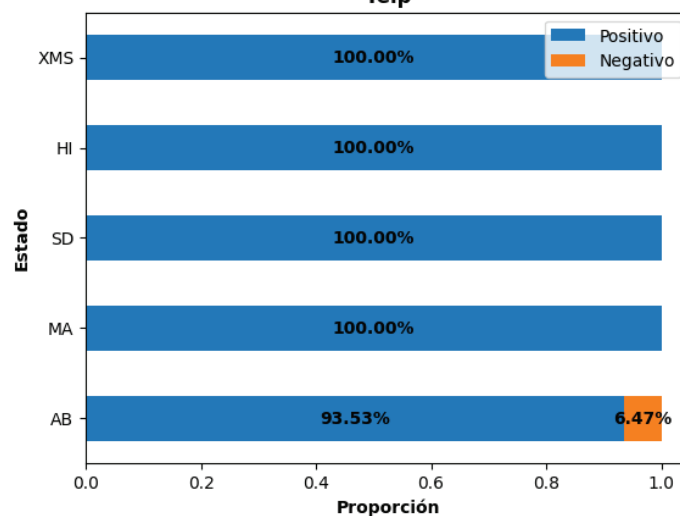
Top 5 Estados con Mayores Porcentajes de Sentimientos Positivos Google Maps Top 5 Estados con Mayores Porcentajes de Sentimientos Negativos Google Maps



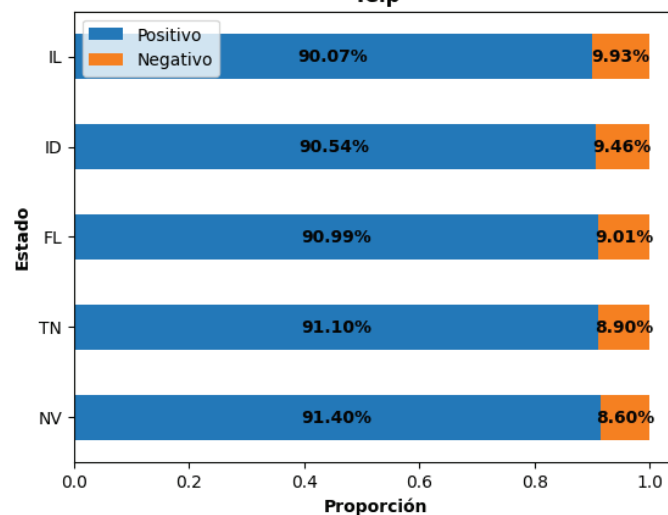
Yelp! - RESULTADO TABLA:

	Nombre_Local	Fecha	Hora	Estado	Servicio	Sentimiento
0	Bloom Salon & Beauty Bar	2020-10-01	17:22:53	XMS	Hair Stylists	Positivo
1	Raed's Hair & All	2019-04-14	11:14:12	MA	Beauty & Spas	Positivo
2	Raed's Hair & All	2016-01-17	22:06:50	MA	Beauty & Spas	Positivo
3	Omkaara Tattoo	2018-07-20	03:11:41	HI	Tattoo	Positivo
4	Absolute Dental	2017-06-20	17:17:51	SD	General Dentistry	Positivo
...
30816	Salon U	2017-12-01	21:19:25	TN	Hair Salons	Positivo
30817	Milano Nails Salon	2019-06-15	00:18:04	TN	Nail Salons	Positivo
30818	Phoebus Tattoos Studio	2018-01-21	21:37:28	TN	Beauty & Spas	Positivo
30819	Thread Bar	2018-07-18	01:23:00	TN	Henna Artists	Positivo
30820	Chlitina	2019-08-14	12:42:17	TN	Skin Care	Positivo

Top 5 Estados con Mayores Porcentajes de Sentimientos Positivos Yelp



Top 5 Estados con Mayores Porcentajes de Sentimientos Negativos Yelp





MACHINE LEARNING CLUSTERING

A través de un modelo de clustering en tres dimensiones (latitud, longitud y promedio de rating) se investigan y se agrupan los negocios. Esto orientado a sus ubicaciones geográficas específicas (Estado y Condado/County) junto a sus tendencias de rating. Complementando así con la vigente competencia en cada locación; ordenado desde el elemento con mas rating hacia el menor. En el mismo se utilizó la librería sklearn donde se extrajo las sublibrerías StandardScaler para estandarizar los datos, KMeans para el proceso de clustering y por ultimo una segunda librería llamada geodesic del conjunto de geoppy para identificar los condados y Estados de las ubicaciones investigadas.

	Promedio_Puntaje_Reviews	Cantidad de Negocios	Negocios Competidores	Estado	Condado	Porcentaje de competición
0	4.96	3459	2447	Pennsylvania	Chester County	70.74
1	4.95	1907	1282	Illinois	Fayette County	67.23
2	4.94	1303	803	Louisiana	Cameron Parish	61.63
3	4.90	2518	1234	North Carolina	Macon County	49.01
4	4.87	1252	765	Oregon	Gilliam County	61.10
5	4.80	2171	1350	Michigan	Van Buren County	62.18
6	4.77	1386	815	Massachusetts	Franklin County	58.80
7	4.75	2582	1334	California	Los Angeles County	51.67
8	4.74	4192	2114	Florida	Hillsborough County	50.43
9	4.73	704	373	Minnesota	Wright County	52.98
10	4.72	1584	727	Arizona	Pinal County	45.90
11	4.71	2128	985	California	El Dorado County	46.29
12	4.70	1532	924	Texas	Grayson County	60.31
13	4.63	979	530	Colorado	Eagle County	54.14
14	4.50	1521	440	Louisiana	Cameron Parish	28.93

Tabla resultante:

Gracias al proceso que implica el clustering al extraer los puntos centrales de cada cluster tenemos una ubicación geográfica exacta de dicho conjunto. Es así como se procede a traducir estos datos en las columnas "Estado" y "Condado"; agregando a su vez el promedio de los puntajes de las reviews de dicha zona. Por otro lado podemos ver a la vez una comparación entre la Cantidad de Negocios presente vs la cantidad de Negocios Competidores, refiriéndose a un rubro en común; la última columna expresa la relación de esta competencia, es decir, cuanto mas alto es el porcentaje, mayor competencia se encuentra presente en esta ubicación. Gracias a este conjunto de datos como parámetros de decisión se puede dar a conocer a cuales Estados y Condados conviene invertir dependiendo del promedio de Rating, como a su vez teniendo en cuenta el porcentaje de competición presente en cada zona. Ya que si el cliente desea un entorno de negocio con poca competencia entonces conviene buscar una zona con poca competencia en comparación a los demás. En caso contrario si el cliente desea sumarse al grupo de competentes en las zonas presentes también es factible ya que simboliza que el negocio da frutos en dicho lugar.

Se puede observar que dicho gráfico indica que el número óptimo de clusters a aplicar son de aproximadamente 5. Pero el contexto de negocio de este proyecto nos exige una clasificación de las ubicaciones con mayores particiones. Por eso se optó por utilizar la cantidad de 50 clusters; así podemos tener 50 localizaciones distintas

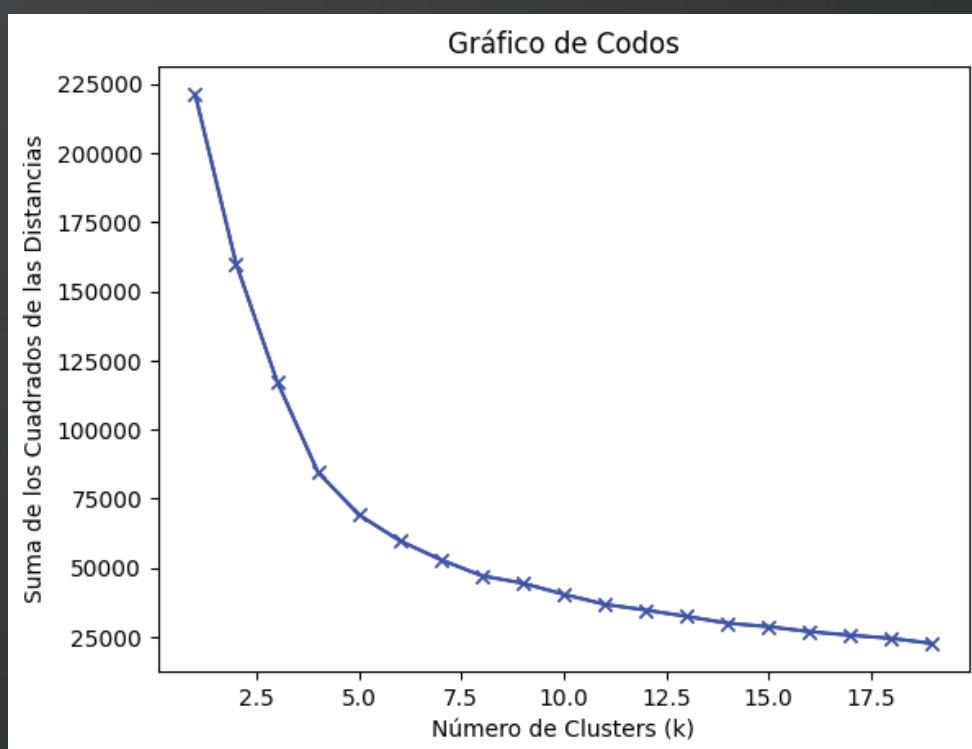
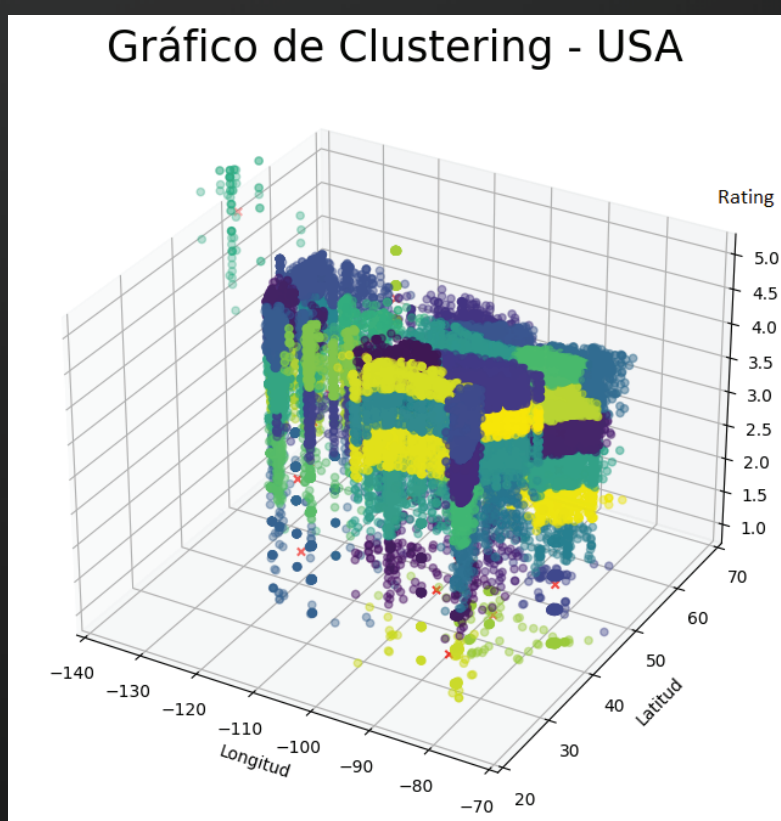


Gráfico 3D del clustering

En este gráfico se puede apreciar inicialmente la forma tridimensional de Estados Unidos, donde los colores representan los clusters conjunto a sus contenidos (puntos).



GRACIAS 🚀

