



DIPLOMADO DE INTELIGENCIA ARTIFICIAL Y
APRENDIZAJE PROFUNDO



Sistema de recomendaciones de compra basado en un modelo de Asignación Latente de Dirichlet

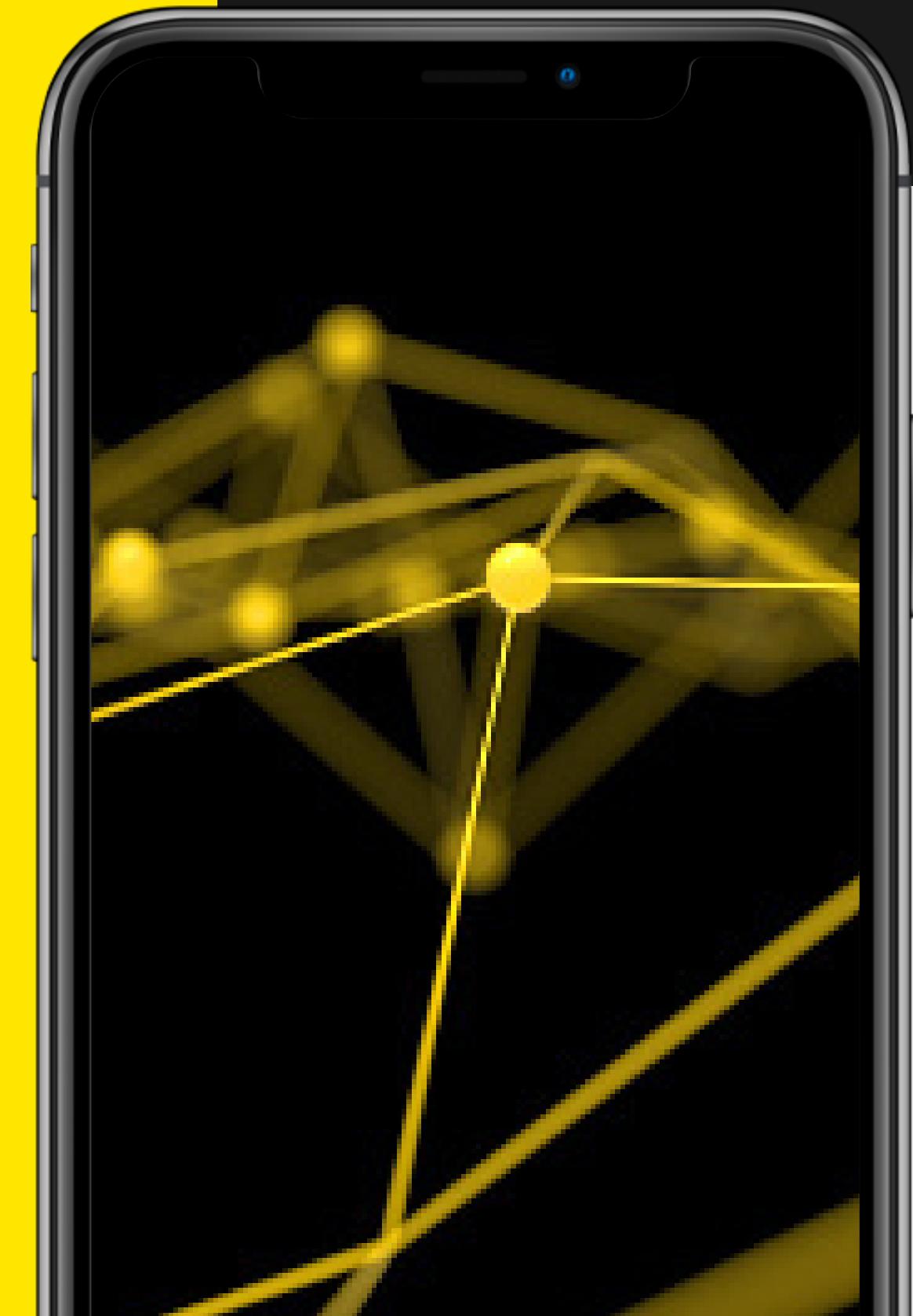


Andrea Valencia Orozco

Andrés Murcia

Diana Carolina Cárdenas

Mayo 2021





Agenda

- Introducción
- Planteamiento del problema
- Solución Propuesta
 - Pre-procesamiento
 - Modelado LDA
- Resultados
- Conclusiones





Planteamiento del problema

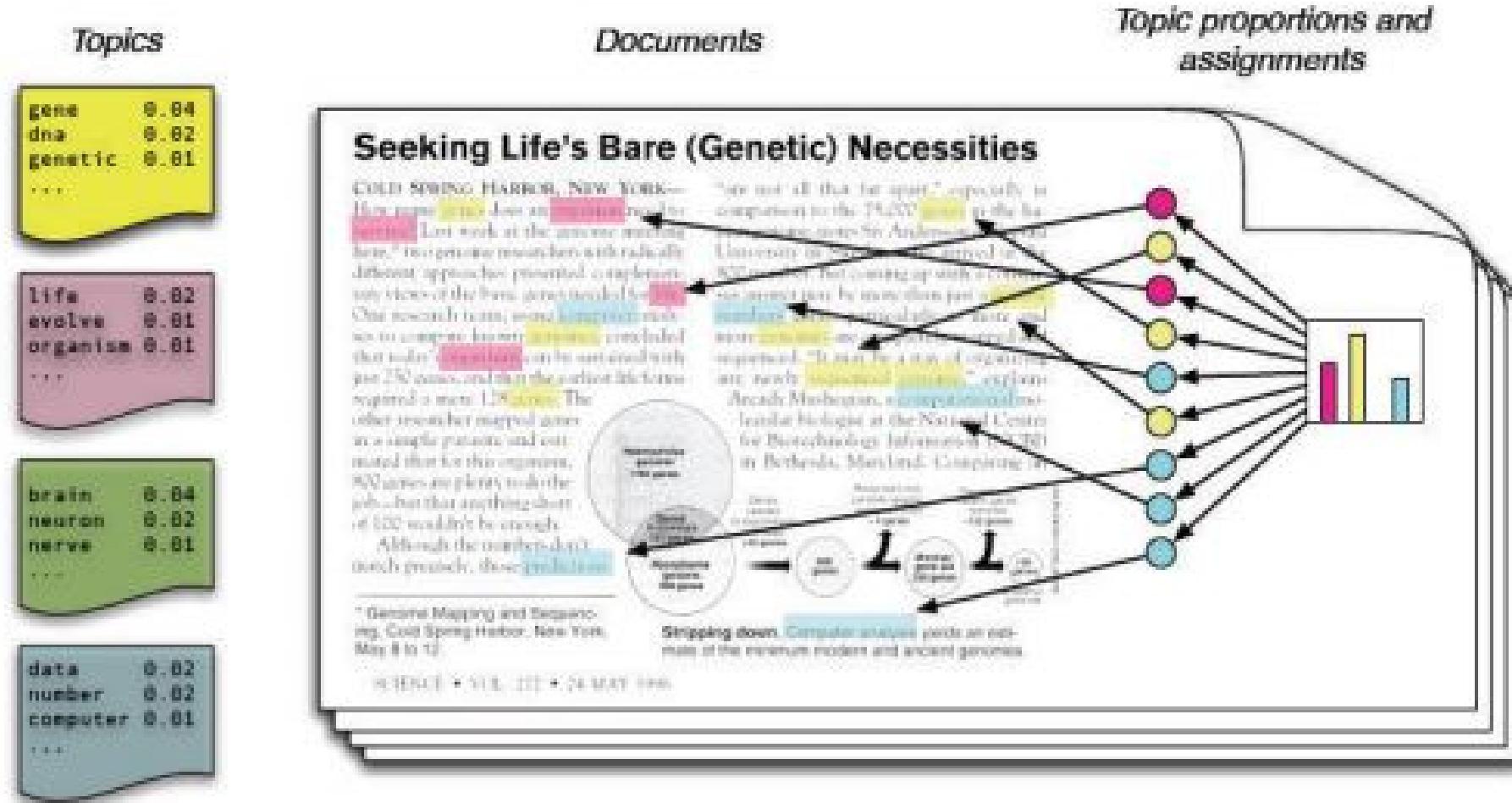
Debido al alto volumen de pedidos, variedad de productos y de establecimientos, la compañía desea anticiparse a los requerimientos de los clientes (tiendas), basado en sus compras de meses anteriores.



Los datos:

- 1'352.921 registros de compra.
- ID Cliente.
- ID Producto.
- Cantidad.
- Fecha.

Solución Propuesta



Fuente: Blei, D., Ng, A., Jordan, M. (2010) *Latent Dirichlet Allocation*

- Las observaciones son generadas por un proceso probabilístico generativo que incluye variables latentes.
- Se infiere la estructura latente respecto a cuáles tópicos describen esa colección de documentos.
- Validación del tópico acertado para nuevos datos probados en el modelo estimado.

Pre-procesamiento →

- Conformación del corpus
- Tokenización
- No eliminación de palabras .

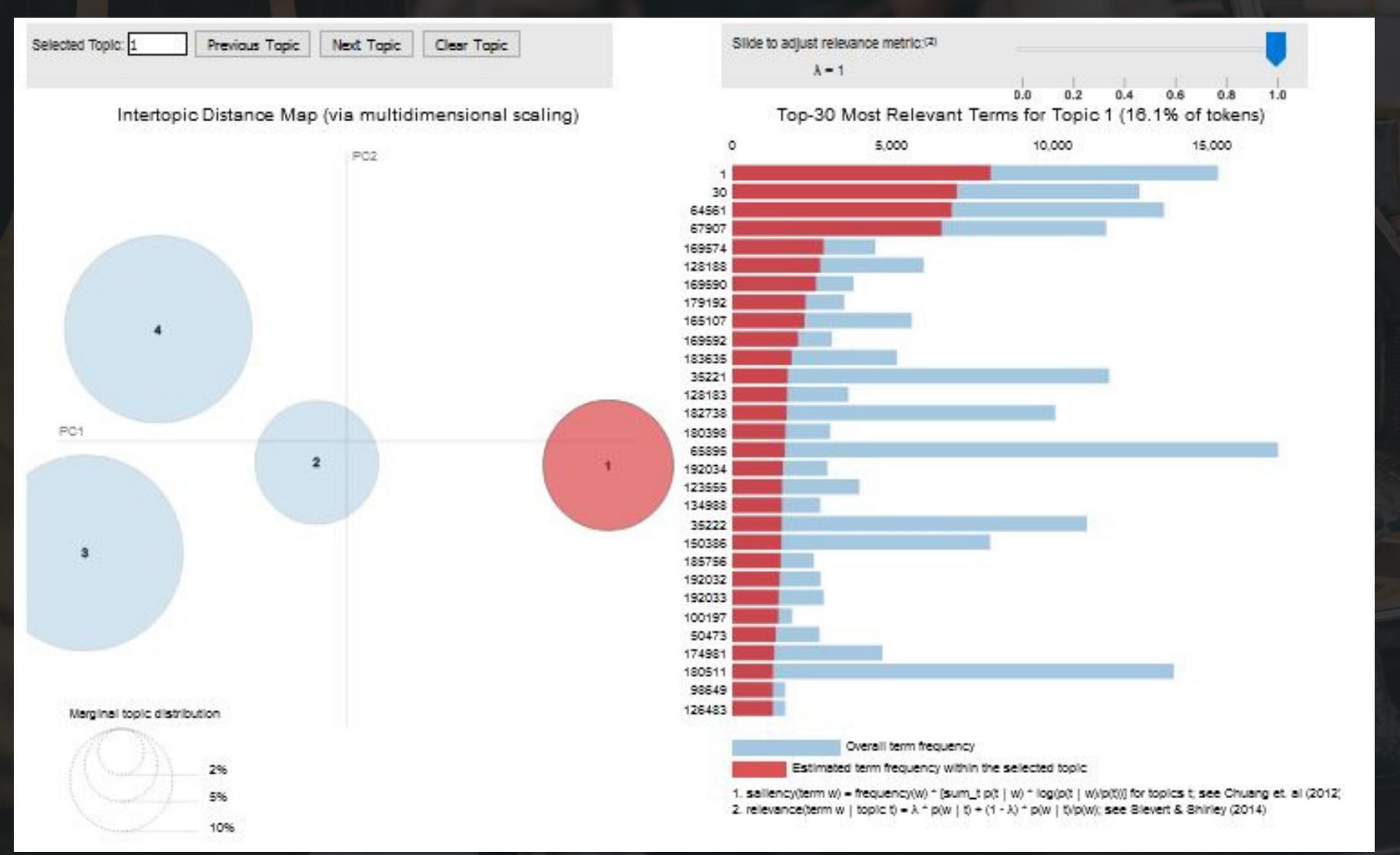
Modelado LDA →

- Selección del número de tópicos.
- Cada tópico como conformación de palabras (productos), cada una con un peso.
- Gráficas de tópicos usando pyLDAvis.

Prueba del modelo →

- Prueba de datos en productos nuevos, con por lo menos una compra.

Resultados Modelamiento



Tópico	Descripción
1	Mayor presencia de cervezas clásicas en presentaciones personales (350 ml) en botella de vidrio, al igual que amplio mix de bebidas de té de sabores surtidos en presentaciones personales (350 ml) en botella de vidrio.
2	Predominancia de cervezas clásicas, además, cervezas bajas en calorías, de presentaciones de máximo 1 litro en botella de vidrio. En menor medida, manejo de algunos items de gaseosas oscuras, energizantes (en packs de 12 unidades) y bebidas de té.
3	Predominancia de gaseosas oscuras de máximo 2.5 litros en botella de plástico. En la misma medida se presentan bebidas de té de sabores surtidos de 2.5 litros en botella de plástico. En menor medida se encuentra asociado a bebidas energizantes (en packs de 12 unidades) de 600 ml.
4	Amplia variedad de cervezas en lata. Presencia de cervezas clásicas de 1 litro en botella de vidrio. En menor medida se presentan bebidas energizantes (en packs de 12 unidades) de contenido de 600 ml.

- Tópico 1: 16.1%; Tópico 2: 14.4%; Tópico 3: 36.3%; Tópico 4: 33.2%.

Resultados Predicciones

Tienda	Tópico	Score	Productos
10	0	0,611	6589, 50469, 180511, 164851, 150428, 35221, 180523, 164478, 63352, 11
236	2	0,736	102316, 11, 64561, 180826, 67907, 16485, 6589, 151754, 1515754, 151753, 50469
150	0	0,746	6589, 50469, 180511, 164851, 150428, 35221, 180523, 164478, 63352, 11
2	0	0,619	6589, 50469, 180511, 164851, 150428, 35221, 180523, 164478, 63352, 11
399	3	0,536	11, 30, 64561, 67907, 128188, 183635, 165107, 169574, 167699, 173742

- El entregable corresponde al tópico y productos recomendados para el nuevo cliente.
- Se sugiere comparar este sistema de recomendaciones con los pedidos reales de los clientes en la siguiente factura de compra.
- Es importante que estos resultados sean enriquecidos con el conocimiento de los expertos en las áreas de negocio de la compañía.

Conclusiones generales

Metodología LDA

Resuelve oportunamente la tarea de detección de tópicos al definir el número óptimo de ellos con el criterio de perplexity.

Pre-procesamiento de los datos

Se deben considerar todos los productos como palabras importantes, a menos que los especialistas de las diferentes áreas de la compañía consideren que uno o varios productos son ofrecidos por defecto a las tiendas.

Tópicos hallados

Se distinguen cuatro (4), que van desde los más inclinados a los productos de bebidas gaseosas y bebidas de té en sus diferentes sabores y envases, hasta los tópicos de mayor afinidad con cervezas de diferentes presentaciones.

Next Steps

Actualizar el modelo cada seis meses. Posibles casos de uso (aplicaciones de recomendaciones).

Equipo de trabajo

Diplomado Inteligencia Artificial y Aprendizaje Profundo
UNAL



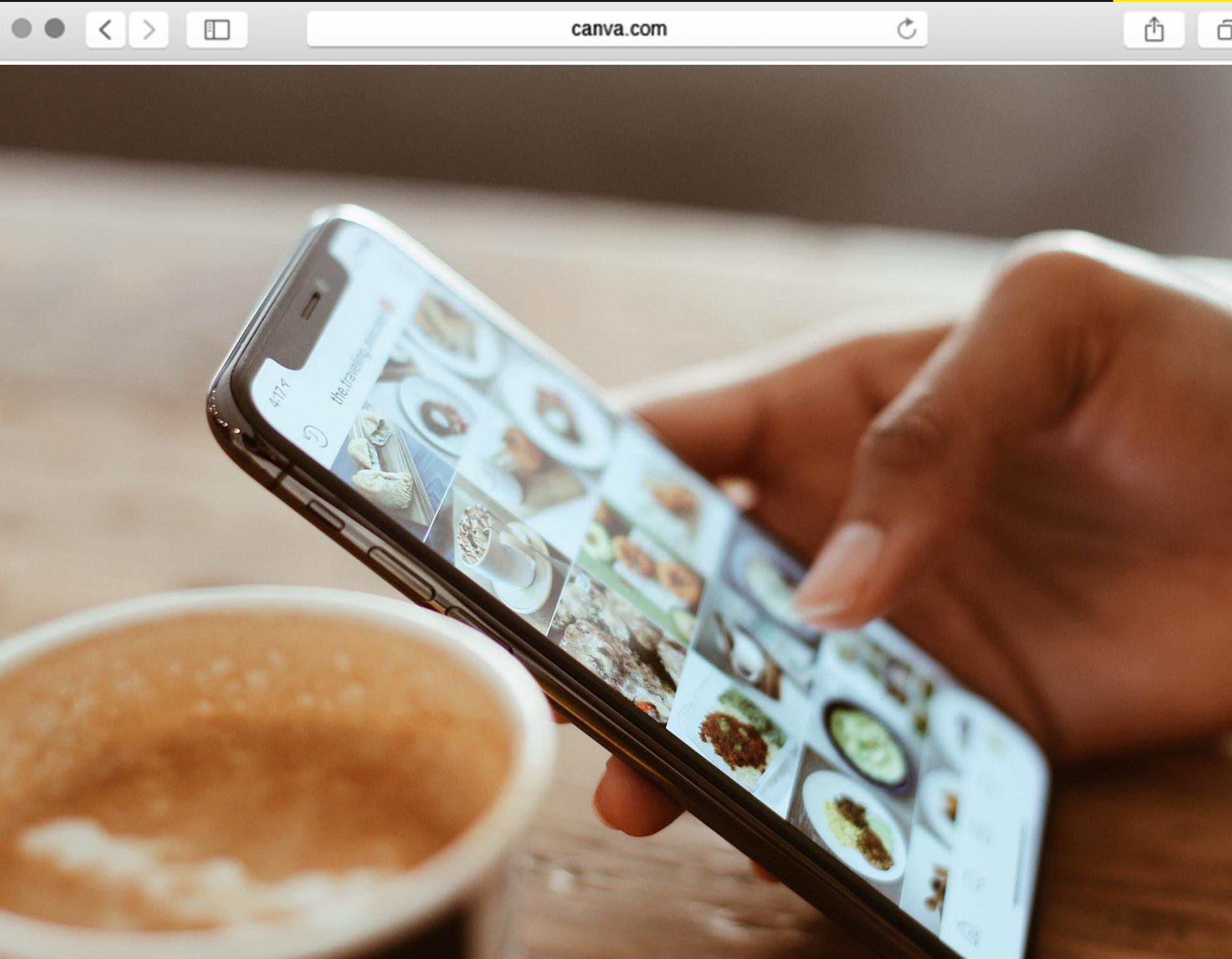
Andrés Murcia
Ingeniero de sistemas



Andrea Valencia
Estadística



Diana Cárdenas
Estadística



GRACIAS