



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

**DESARROLLO DE UNA HERRAMIENTA DE VISUALIZACIÓN
INTERACTIVA PARA CARACTERIZAR LA CALIDAD NUTRICIONAL DE
RESTAURANTES**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN.

GERMÁN GUSTAVO IGNACIO URREA MANSILLA

PROFESOR GUÍA:
Andrés Abeliuk Kimelman

MIEMBROS DE LA COMISIÓN:
Francisco Gutiérrez Figueroa
José Urzúa Reinoso

SANTIAGO DE CHILE
2024

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN
POR: GERMÁN GUSTAVO IGNACIO URREA MANSILLA
FECHA: 2024
PROF. GUÍA: Andrés Abeliuk Kimelman

DESARROLLO DE UNA HERRAMIENTA DE VISUALIZACIÓN INTERACTIVA PARA CARACTERIZAR LA CALIDAD NUTRICIONAL DE RESTAURANTES

La alimentación es una de las necesidades básicas más importantes de cualquier ser. Si bien una alimentación balanceada puede ser increíblemente beneficiosa para la salud física y mental, una alimentación de calidad deficiente puede desencadenar en problemas de salud como osteoporosis, obesidad y diabetes de tipo 2. En este contexto, uno de los factores de mayor interés al momento de estudiar la calidad de la alimentación en una población es el entorno alimentario, el cual, ya sea por causas económicas o sociales, puede dificultar el acceso a alimentos de un valor nutricional adecuado. Debido a la importancia del entorno, han surgido múltiples iniciativas y estudios para comprender mejor estos y estudiar posibles formas de influir en estos entornos. En este contexto, un equipo de investigación de la Universidad del Sur de California (USC), financiado por el Centro de Salud para Latinos de California del Sur, ha desarrollado un indicador de densidad nutricional en base a modelos de lenguaje capaces de predecir la composición nutricional de alimentos en base a su nombre.

Es en base al modelo desarrollado, que se busca crear un prototipo de portal web capaz de presentar los indicadores derivados del modelo desarrollado. En particular, el objetivo principal de este trabajo de título es desarrollar una aplicación capaz de generar conocimiento sobre la calidad y desigualdad nutricional en la ciudad de Los Angeles, en Estados Unidos. Para ello se desarrollan visualizaciones, como un mapa interactivo que presenta la densidad nutricional de distintas áreas administrativas y restaurantes en el estado de California, así como visualizaciones estadísticas que correlacionan el indicador de densidad nutricional a nivel de entorno (FEND) con otras variables de interés, como indicadores de prevalencia de enfermedades relacionadas a una alimentación de baja calidad.

El prototipo desarrollado, evaluado por la investigadora Abigail Horn de la USC, logró cumplir con las expectativas a nivel de funcionalidad. Las visualizaciones fueron consideradas útiles para detectar zonas de alta y baja densidad nutricional, así como para poder intuir posibles relaciones entre variables de interés y el indicador FEND. A pesar de la recepción favorable a nivel funcional, existieron comentarios constructivos respecto a la retroalimentación para el usuario, en donde aspectos como la descarga de archivos o el uso del mapa interactivo podrían haber sido estructurados, explicados o guiados de una forma más amigable para un usuario nuevo.

Se concluye que el prototipo es una buena base para futuras iteraciones. En donde sería relevante primero abordar aspectos a nivel estético y de usabilidad, con el objetivo de mejorar la experiencia de un usuario nuevo, y posteriormente expandir las funcionalidades de la aplicación para que un usuario cualquiera pueda interactuar directamente con el modelo de lenguaje desarrollado.

*You may have lost, but life goes on. So, please, don't ever think,
"I'm such a failure", or anything of the sort.*

Kazuki Takahashi

Agradecimientos

A todas las personas que han pasado por mi vida, ya sea que me hayan apoyado o criticado, todos me ayudaron a crecer y llegar hasta aquí, para bien o para mal.

Agradezco en especial a mi familia y amigos, por el constante apoyo y recordarme que la vida no sólo tiene bajos, sino también altos. Agradezco también a mi profesor guía y el equipo de la USC, sin su apoyo y retroalimentación este documento no se habría completado.

Tabla de Contenido

1. Introducción	1
1.1. Contexto	1
1.2. Comprendiendo mejor los entornos alimentarios	2
1.3. Situación actual	2
1.4. Objetivos	3
1.4.1. Objetivos Específicos	3
2. Marco Teórico	4
2.1. Tecnologías y arquitecturas de desarrollo	4
2.1.1. Arquitectura REST	4
2.1.2. API REST	5
2.1.3. Django REST Framework	5
2.1.4. ReactJS	5
2.1.5. Bases de datos espaciales	5
2.2. Indicadores	6
2.2.1. Métrica RRR	6
2.2.2. Métricas derivadas del RRR	6
3. Diseño de la solución	7
3.1. Exploración de datos	7
3.1.1. Predicciones del indicador nutricional	7
3.1.2. Datos geográficos	8
3.1.3. Indicadores de salud	8
3.1.4. Otros indicadores	9
3.2. Construcción del modelo de datos	9
3.2.1. Entidades derivadas de las predicciones del RRR	9
3.2.2. Entidades derivadas de datos geográficos	10
3.2.3. Entidades derivadas de indicadores	10
3.2.4. Modelo de datos final	11
3.3. Diseño del portal web	12
3.3.1. Requisitos de usuario	12
3.3.2. Arquitectura física	13
3.3.3. Arquitectura lógica	13
3.3.4. Diseño inicial de la aplicación	14
3.3.4.1. Navegación y mapa interactivo	14
3.3.4.2. Visualización de estadísticas	14
3.3.4.3. Descarga de datos	15

3.3.4.4.	Información adicional	16
3.3.5.	Tecnologías	16
3.3.5.1.	Backend y Base de datos	16
3.3.5.2.	Frontend y visualizaciones	16
4.	Implementación de la solución	18
4.1.	Base de datos	18
4.1.1.	Preprocesamiento de datos	18
4.1.2.	Población de base de datos	18
4.2.	Backend	19
4.3.	Interfaces	19
4.3.1.	Mapa interactivo	19
4.3.2.	Estadísticas	24
4.3.3.	Descargas	25
4.3.4.	Información	26
5.	Validación	27
5.1.	Pruebas de validación	27
5.1.1.	Tareas ejecutadas	27
5.1.1.1.	Tareas en sección inicial	27
5.1.1.2.	Tareas en sección de estadísticas	28
5.1.1.3.	Tareas en sección de descargas	28
5.1.1.4.	Tareas en sección de información	28
5.1.2.	Preguntas de validación	28
5.1.2.1.	Preguntas enfocadas a investigadores	29
5.1.2.2.	Preguntas para usuarios ajenos a la investigación	29
5.2.	Primera iteración	29
5.3.	Segunda iteración	30
5.3.1.	Mapa interactivo	30
5.3.2.	Visualizaciones estadísticas	31
5.3.3.	Carga y descarga de datos	31
5.3.4.	Posibles mejoras	31
5.4.	Discusión	32
5.4.1.	Cumplimiento de objetivos	32
5.4.2.	Mejoras de optimización	32
5.4.3.	Apreciación personal	32
6.	Conclusión	35
6.1.	Cumplimiento de objetivos	35
6.2.	Trabajo futuro	36
	Bibliografía	37
	Anexos	39
A.	Preguntas sobre usabilidad y eficiencia.	39

Índice de Ilustraciones

3.1.	Modelo de datos	12
3.2.	Arquitectura simple de cliente-servidor.	13
3.3.	Mock-up para vista de mapa interactivo	14
3.4.	Mock-up de idea inicial para vista de visualización de densidad nutricional en base a características de interés.	15
3.5.	Mock-up de idea inicial para vista de descarga de datos.	15
3.6.	Mock-up de ejemplo para sección de información adicional.	16
4.1.	Vista inicial del mapa interactivo	20
4.2.	Ejemplo de popup en mapa interactivo	20
4.3.	Ejemplo de detalles en barra lateral	21
4.4.	Ejemplo de detalles en barra lateral, cuando no es posible procesar la cantidad de menús	22
4.5.	Controlador del mapa	23
4.6.	Mapa interactivo con indicador por restaurantes, mostrando solamente los restaurantes de cadena.	23
4.7.	Mapa interactivo con indicador por ambientes, a nivel <i>places</i> , filtrando las zonas de baja densidad nutricional.	23
4.8.	Vista por defecto de las estadísticas, junto a la selección de variables.	24
4.9.	Correlación entre el FEND y la prevalencia de presión alta en adultos, a nivel de lugares (<i>places</i>).	24
4.10.	Correlación entre el FEND y la prevalencia de presión alta en adultos, a nivel de lugares (<i>places</i>).	25
4.11.	Correlación entre el FEND y la prevalencia de presión alta en adultos, a nivel de lugares (<i>places</i>).	25
4.12.	Vista de descarga de datos	26
4.13.	Vista de sección de información	26
5.1.	Vista por defecto del mapa, en vista simulada de un Samsung Galaxy S8 (Resolución 360 × 740)	33

Capítulo 1

Introducción

1.1. Contexto

La alimentación es una parte fundamental en el día a día de cualquier ser humano, es una necesidad básica por naturaleza. Es bien sabido que una alimentación balanceada con un alto valor nutricional puede ser vital en la calidad de vida de cualquier persona, ayudando a las funciones digestivas, la mantención de un peso saludable, y el sentimiento de bienestar general. Sin embargo, una alimentación de calidad deficiente, como una alta en grasas saturadas y azúcares procesados, puede desencadenar en problemas de colesterol alto, presión alta, e incluso enfermedades como la osteoporosis, obesidad y diabetes tipo 2. En particular, se sabe que para el año 2021 la diabetes y enfermedades crónicas tanto al hígado como al corazón estuvieron entre las 10 causas más comunes de muerte en los Estados Unidos de América [22].

Esta problemática se ve agravada por el entorno alimentario, el cual, ya sea por causas económicas o sociales, dificulta el acceso a comida de calidad, o, por el contrario, facilita el acceso a comida rápida y/o de baja calidad nutricional. El entorno alimentario es un factor importante a considerar, pues se sabe que las comunidades compuestas por personas de origen hispano, afroamericano y/o de bajos ingresos están expuestas a entornos alimentarios de peor calidad, siendo un caso particular las comunidades latinas con residencia en los Estados Unidos [18].

Es debido a la importancia del entorno alimentario que han surgido programas para ejercer influencia en estos. Un ejemplo a mencionar sobre este tipo de programas es la “Iniciativa de financiamiento a la alimentación saludable” [6], una iniciativa por parte de la organización privada sin ánimos de lucro “Reinvestment Fund” [9], esta iniciativa financia y ofrece asistencia a proyectos localizados en Estados Unidos cuyo objetivo sea facilitar el acceso a comida saludable en zonas rurales y urbanas de bajos recursos. Aun así, iniciativas como la mencionada no han demostrado mejoras sustanciales en la calidad de alimentación o en la disminución de enfermedades relacionadas con la nutrición [14]. Una limitante en este tipo de iniciativas es la falta de habilidad para medir y caracterizar correctamente los entornos alimentarios, impidiendo de esta forma evaluar rigurosamente los cambios necesarios para la mejora de dichos entornos, así como sus efectos concretos en la salud nutricional.

1.2. Comprendiendo mejor los entornos alimentarios

El Centro de Salud para Latinos de California del Sur, también conocido como “The Southern California Center for Latino Health” (SCCLH), es una institución asociada al hospital de niños de Los Ángeles cuyo objetivo es mejorar la calidad de vida de las familias latinas en Estados Unidos mediante la investigación de las posibles disparidades en la salud que afecten a las comunidades latinas, y el posterior desarrollo de soluciones a estas [10]. En el contexto de la comprensión de los entornos alimentarios, el SCCLH financia el desarrollo de una herramienta para caracterizar la calidad nutricional de los restaurantes, liderada por un equipo de investigación conformado por estudiantes y académicos de la “University of Southern California” (USC), desarrollando un indicador para evaluar la densidad nutricional de un entorno. Posteriormente, los resultados obtenidos se utilizan para obtener información sobre las posibles disparidades nutricionales que puedan existir entre los entornos alimentarios de los vecindarios de Los Ángeles, en particular aquellos con predominancia latina en comparación con aquellos que no la tienen.

El desarrollo de esta herramienta se basa en el cumplimiento de tres hitos principales.

El primer hito consiste en mejorar un algoritmo de puntuación nutricional de restaurantes previamente entrenado como prueba de concepto [20], así como un algoritmo de lenguaje natural que relaciona menús de restaurantes con los alimentos genéricos más parecidos, cuya composición nutricional sea conocida. Con este objetivo se ha procesado una mayor cantidad de datos en comparación a la utilizada en un inicio, de esta forma se ha creado una base de datos sobre la composición genérica de distintos alimentos, la cual se ha utilizado para entrenar el algoritmo de estimación de composición nutricional, el cual además se ha estado refinando mediante la aplicación de modelos de lenguaje del estado del arte, como BERT [15]. Por su parte, el algoritmo de puntuación nutricional se ha estado refinando con ayuda de expertos en nutrición, en particular se han evaluado variables como los micronutrientes relevantes a considerar.

El segundo hito consiste en utilizar los algoritmos previamente mencionados para cuantificar de forma más precisa la calidad nutricional de los entornos alimentarios de vecindarios, incluyendo diferencias y disparidades en vecindarios latinos.

El tercer hito consiste en desarrollar un portal web en donde se exponga de forma clara la información obtenida en el segundo hito y se dé acceso a los indicadores de calidad nutricional. Con el cumplimiento de este último hito no solo se le da visibilidad al trabajo desarrollado, sino que además los datos podrán ser utilizados en posteriores estudios de salud nutricional.

1.3. Situación actual

Actualmente, un equipo interdisciplinario de especialistas en el área de la ciencia de datos, modelamiento de sistemas alimenticios, nutrición y disparidades en la salud, conformado principalmente por investigadores de la USC, han realizado avances en el refinamiento del algoritmo de composición nutricional y el cálculo del índice de densidad nutricional. En par-

ticular, los resultados obtenidos ya son aceptables para generar muestras de los datos.

Es en este contexto, que el equipo de investigación ha considerado viable comenzar el desarrollo del prototipo mencionado en el tercer hito. Del cual no se posee ningún desarrollo hasta el momento.

1.4. Objetivos

El presente trabajo de título se enmarca en el tercer hito presentado. El objetivo principal de esta memoria es implementar un prototipo de portal web que permita generar conocimiento sobre la calidad y desigualdad nutricional en distintas zonas de una misma ciudad o comuna, en el caso particular de esta memoria se abarcaría solamente la ciudad de Los Ángeles y zonas cercanas en el estado de California. La idea es que el prototipo desarrollado sea capaz de entregar información de interés a investigadores del área de la salud y/o de las ciencias sociales, a través de visualizaciones interactivas e indicadores nutricionales, así como la capacidad de obtener los datos para sus propias investigaciones.

1.4.1. Objetivos Específicos

Para lograr el objetivo principal se deben lograr una serie de objetivos más pequeños, los cuales se listan a continuación.

1. Generar un mapa interactivo que incluya información útil sobre la distribución de los indicadores en áreas administrativas seleccionadas por el usuario.
2. Generar visualizaciones que permitan conocer la distribución de los indicadores en forma general, y su relación con variables de interés, como lo puede ser la permanencia de problemas a la salud.
3. Permitir la descarga de muestras de los datos, de forma que el usuario tenga acceso a los indicadores actualizados para su uso personal.

Capítulo 2

Marco Teórico

En este capítulo se explicarán brevemente las tecnologías e indicadores utilizados durante el desarrollo del presente trabajo de título, con el objetivo de proveer contexto suficiente para comprender los próximos capítulos.

2.1. Tecnologías y arquitecturas de desarrollo

2.1.1. Arquitectura REST

La arquitectura REST (*REpresentational State Transfer*) [12], es un estilo de arquitectura ampliamente utilizado en el desarrollo de servicios web. A continuación se enlistan los principios que sigue la arquitectura REST:

- Arquitectura cliente-servidor compuesta de clientes, servidores y recursos.
- Comunicación entre el cliente y el servidor sin estado, lo cual implica que no se almacena la información del cliente entre las solicitudes.
- Datos que pueden almacenarse en caché y optimizan las interacciones entre el cliente y el servidor.
- Una interfaz uniforme entre los elementos, para que la información se transfiera de forma estandarizada. Para ello deben cumplirse las siguientes condiciones:
 1. Los recursos solicitados deben ser identificables e independientes de las representaciones enviadas al cliente.
 2. El cliente debe poder manipular los recursos a través de la representación que recibe, ya que esta contiene suficiente información para permitirlo.
 3. Los mensajes autodescriptivos que se envíen al cliente deben contener la información necesaria para describir cómo debe procesarla.
 4. Debe contener hipertexto o hipermedios, lo cual significa que cuando el cliente acceda a algún recurso, debe poder utilizar hipervínculos para buscar las demás acciones que se encuentren disponibles en ese momento.
- Un sistema en capas que organiza en jerarquías invisibles para el cliente cada uno de los servidores (los encargados de la seguridad, del equilibrio de carga, etc.) que participan en la recuperación de la información solicitada.

- Código disponible según se solicite (opcional), es decir, la capacidad para enviar códigos ejecutables del servidor al cliente cuando se requiera, lo cual amplía las funciones del cliente.

2.1.2. API REST

Una API REST es una interfaz de programación de aplicaciones (API) que se ajusta a los límites de la arquitectura REST. Las API permiten interactuar con una computadora o un sistema para obtener datos o ejecutar una función, de manera que el sistema comprenda la solicitud y la cumpla [13].

2.1.3. Django REST Framework

Django [3] es un framework de desarrollo web Fullstack escrito en Python el cual sigue el patrón de diseño MVT (Model View Template), sin embargo, para el trabajo de título presentado es de mayor interés las funcionalidades de Backend de este framework. En particular, Django facilita la conexión y manejo de bases de datos mediante su esquema ORM (Object Relational Model), el cual permite tratar con entidades como si fueran objetos de python.

En el contexto de las funcionalidades de backend, Django REST Framework [4] es una extensión que facilita la creación de APIs REST utilizando el modelo ORM de Django. El uso de Django REST permite conservar el uso de las funcionalidades de backend de Django, al mismo tiempo que otorga la flexibilidad de utilizar tecnologías distintas para la creación de la vista de usuario.

2.1.4. ReactJS

ReactJS [8] es una popular biblioteca de Javascript que facilita la implementación de vistas de usuario mediante un patrón de diseño basado en la creación de componentes. En este esquema, cada componente maneja su propia lógica interna y solamente es renderizada cuando es necesario, lo que permite un diseño de aplicaciones modular, ordenado y optimizado. Es debido a esto que es común que React reciba soporte de la comunidad, en forma de bibliotecas de componentes de código abierto, las cuales permiten la adición de funcionalidades sin la necesidad de desarrollar una componente desde cero.

Dado que React es una biblioteca de Javascript orientada a la implementación de vistas de usuario, esta requiere del uso de *requests* a APIs Web con el objetivo de obtener los datos necesarios de la base de datos.

2.1.5. Bases de datos espaciales

Dado que el desarrollo del prototipo involucra el procesamiento de datos geoespaciales, es importante tener en cuenta los métodos utilizados para tratar con este tipo de datos.

En el caso del almacenamiento de este tipo de datos, es común el uso de bases de datos espaciales, bases de datos extendidas para la representación de datos espaciales como objetos geométricos (puntos, líneas y polígonos), incluyendo funcionalidades de consulta y análisis optimizadas para este tipo de datos. Un ejemplo de un sistema de bases de datos espacial

popular es *PostGis* [7], una extensión gratuita del motor de bases de datos *PostgresSQL*, que permite almacenar y realizar operaciones espaciales entre datos.

2.2. Indicadores

2.2.1. Métrica RRR

El “Ratio of Recommended to Restricted nutrients” [16], conocido también como RRR, es la métrica predicha por el modelo de predicción de densidad nutricional. Es una métrica basada en la composición nutricional de un alimento, en particular, se calcula como el ratio entre la densidad de nutrientes “positivos” (proteína, calcio, fibra, etc) y “negativos”(colesterol, azúcar, sodio, etc). Dado que el RRR es un ratio, es invariante al tamaño de la porción de un alimento, y se interpreta que alimentos de mejor calidad son aquellos que superen el valor 1.0.

2.2.2. Métricas derivadas del RRR

Si bien con el modelo de predicción es posible obtener el RRR, esta métrica se limita a la composición nutricional de alimentos. Es debido a esto que para poder estudiar la densidad nutricional a un nivel más general se definen dos métricas de agregación derivadas del RRR.

- **Restaurant Nutrient Density:** También conocida como RND, se describe como la mediana del RRR de todos los menús de un restaurante.
- **Food Environment Nutrient Density:** También conocida como FEND, se describe como la mediana del RND de todos los restaurantes de un área.

Capítulo 3

Diseño de la solución

Dado que el objetivo es implementar un prototipo de portal web que permita visualizar y analizar los indicadores diseñados, el diseño de la solución se enfoca en ir identificando todas las partes necesarias para el correcto funcionamiento del prototipo, de forma que este sea un producto mínimo viable capaz de satisfacer los requerimientos iniciales del prototipo.

3.1. Exploración de datos

Previo a iniciar cualquier desarrollo, es importante primero identificar y analizar los datos disponibles para el desarrollo de la plataforma, dado que el objetivo de la plataforma es poder entregar visualizaciones adecuadas de estos.

El primer paso para poder concebir una solución es identificar las variables relevantes que necesita el portal para funcionar. Esto puede ir desde indicadores nutricionales, de salud, de ingreso, datos geográficos, etc. Estas variables están ligadas principalmente a los datos crudos, ya sea aquellos que se pueden obtener de sitios públicos, o aquellos que son provistos por el equipo de investigación.

3.1.1. Predicciones del indicador nutricional

Como base, se tienen predicciones del score nutricional RRR por menú, para varios restaurantes, estos datos son otorgados por investigadores de la “University of Southern California”. En particular, estas predicciones se realizan a partir de los menús obtenidos por 2 fuentes

- Brizo: Datos de menús para, aproximadamente, 35000 restaurantes en la ciudad de Los Angeles y alrededores. Se cuenta con el nombre del menú, un ID para el establecimiento y su ubicación aproximada.
- Spoonacular: Datos de menús para alrededor de 500 restaurantes, pertenecientes a cadenas de restaurantes, en la ciudad de Los Angeles y alrededores. Se cuenta con el nombre del menú, un ID para la cadena de restaurante de cada establecimiento, así como el nombre de la cadena y ubicación aproximada de cada establecimiento.

3.1.2. Datos geográficos

Siendo el enfoque el poder visualizar la calidad nutricional de ciertas áreas en la ciudad de Los Angeles y alrededores, es necesario contar con datos geográficos de, al menos, el condado de California. Para ello se cuenta con datos públicos de la página oficial del censo estadounidense, *Census.gov* [1], estos datos contienen los límites de múltiples tipos de distribuciones administrativas a lo largo de Estados Unidos. Para el presente trabajo de título se escogieron 3 tipos de distribuciones sobre las cual evaluar el indicador de densidad nutricional de entorno (FEND):

- **County:** Conocido en español como condado. Es la unidad más grande de distribución administrativa y política en un estado.
- **Census Tract:** Conocido en español como zona censal. Es la siguiente unidad administrativa después del condado, son zonas pequeñas de subdivisión delimitadas con el objetivo de tomar el censo, esta es una de las entidades territoriales más pequeñas de la cual se cuenta con datos demográficos, coincidiendo con los límites de ciudades y pueblos. Así, por ejemplo, cada ciudad está compuesta por varias zonas censales.
- **Place:** Conocido en español como lugar. A diferencia de los condados y zonas censales, estos no son necesariamente un subconjunto de otra unidad administrativa, sino que más bien son ubicaciones reconocidas de forma oficial. Ejemplos de lugares son la ciudad de Palmdale, o la misma ciudad de Los Ángeles.

Si bien estos 3 niveles administrativos pueden tener distintos atributos, se consideran solo los comunes para la posterior construcción del modelo de datos. En particular, los campos de interés incluyen los IDs geográfico de 14 y 7 dígitos, nombre, y el polígono que define los límites geográficos.

La elección de estos 3 tipos de distribuciones administrativas se deben principalmente a que cada una permitiría identificar la distribución nutricional en una escala distinta. En particular, los condados, al ser la unidad más grande dentro de un estado, permitirían conocer un panorama más general para la distribución nutricional. Las zonas censales, al ser el siguiente nivel de distribución después de los condados, y una unidad considerablemente más pequeña, permitiría conocer la distribución en regiones más específicas. Por último, los lugares se toman en cuenta como una forma más intuitiva de identificar la distribución nutricional, dado que los lugares suelen tener nombres fácilmente reconocibles y son de conocimiento popular, a diferencia de las zonas censales.

3.1.3. Indicadores de salud

Contando con el indicador principal y la información geográfica de las unidades administrativas, solo falta contar con indicadores de salud relevantes para correlacionar con el indicador nutricional principal. En este caso, la fuente principal de indicadores de salud es el portal de datos del “Center for Disease Control and Prevention” [2], en particular, se utiliza la versión 2022 de PLACES, una base de datos con distintos indicadores de salud en hasta 4 niveles administrativos (*places, counties, census tracts, y ZIP Code Tabulation Areas*). Los datos se descargan por nivel administrativo, sin embargo, todos cuentan con una estructura de datos similar, incluyendo campos como:

- Identificador geográfico de 7 dígitos, mismo que el utilizado por el censo.
- Valor del indicador
- Nombre del indicador (Por ejemplo, obesidad en mayores de 18)
- Identificador del indicador
- Categoría del indicador (Salud, prevención, etc)
- Unidad de medida (Por ejemplo, porcentaje)
- Tipo de dato (Por ejemplo, prevalencia)
- Identificador del tipo de dato

3.1.4. Otros indicadores

Adicionalmente, por sugerencia de la investigadora de la USC Abigail Horn, se añaden algunos indicadores económicos y demográficos. En cuanto a los datos económicos, estos constituyen principalmente el ingreso medio por hogar y algunas variables categóricas sobre la situación relativa del ingreso en una zona censal, estos datos se obtienen de “Los Angeles GeoHub” [19]. Por otro lado, se añaden algunas variables relacionadas con el acceso a fuentes de alimentos en zonas censales, estas se obtienen del “Food Research Atlas” [5] de la USDA (United States Department of Agriculture).

3.2. Construcción del modelo de datos

Una vez identificadas las variables relevantes y analizada la estructura de datos, el siguiente paso es construir un modelo de datos adecuado para la transferencia eficiente de estos hacia el portal web. Este es un proceso que requiere primero del diseño del modelo de datos, posteriormente la elección de un motor de bases de datos adecuado, el procesamiento de los datos crudos y posterior subida de los datos crudos a la base de datos, y por último la generación o cruce de variables relevantes y/o convenientes para el portal web.

En este capítulo el enfoque es el diseño del modelo, en particular, se reconocen 3 tipos de datos: Predicciones del RRR, datos geográficos e indicadores. Tomando en cuenta estos 3 tipos de datos, se puede diseñar una base de datos acorde.

3.2.1. Entidades derivadas de las predicciones del RRR

Partiendo por las predicciones del RRR, se considera prudente separar los menús de los restaurantes, esto con el objetivo de poder entregar la información general del restaurante sin la necesidad de iterar sobre los menús de este. Dado el supuesto de que todos los restaurantes de cadena proveen el mismo menú, o en otras palabras, todos aportan el mismo valor del índice por restaurante (RND), es posible separar los datos en 3 entidades:

- **Menús:** Estos son los datos crudos de los menús, con el nombre del menú, su correspondiente predicción de RRR y un identificador correspondiente al restaurante al que pertenece.
- **Restaurantes:** En este contexto, se considera un restaurante como una entidad con un menú particular, o sea, una cadena de restaurantes solo tendría una entrada en esta tabla. Atributos de esta tabla incluyen el indicador RND, estadísticas de distribución del RRR del restaurante (máximo, mínimo, desviación estándar), nombre, y si es cadena o no.
- **Instalaciones:** Corresponden a todas las instalaciones conocidas de restaurantes. Sus atributos incluyen la ubicación geográfica de la instalación, el identificador del restaurante al que pertenecen, así como nombre y dirección de la instalación.

Es de notar que, si bien en la tabla de restaurantes se cuentan con estadísticas distintas al RRR, estas no vienen en los datos fuente, por lo que son campos de agregación generados posteriormente. La razón de que estos campos de agregación sean almacenados, se debe a que permite agilizar las consultas de estos campos, ya que son requeridos directamente por el mapa, y dado que los datos no deberían actualizarse con frecuencia, esto no debería suponer un problema mientras se ejecuten las consultas SQL correspondientes, por ejemplo, añadiendo un *trigger* que se ejecute al añadir nuevos datos, aunque esto no se implementó debido a que los datos fueron fijos durante el desarrollo.

3.2.2. Entidades derivadas de datos geográficos

En este caso el diseño se hereda directamente de la estructura de los datos crudos. Si bien se puede optar por una estructura jerárquica en base a los niveles administrativos, en donde cada nivel administrativo es una entidad que se relaciona con otro nivel administrativo, se considera mucho más conveniente juntar todos los tipos de geografías en una sola entidad, dado que en este caso lo que interesa es poder juntar de forma sencilla los datos de indicadores junto a los datos geográficos, y para este propósito conviene conservar todos los datos geográficos en una sola tabla en la que se junten los campos comunes de cada nivel administrativo, de forma que el cruce de datos sea directo. De esta forma se tiene la entidad *Geografías* contando con los siguientes atributos:

- Identificadores geográficos global (14 dígitos) y local (últimos 7 dígitos del identificador global).
- Nombre simple de la geografía, y nombre incluyendo descripción a nivel estadístico/legal (lsad). Por ejemplo, el nombre simple de una zona censal puede ser “55.7”, mientras que su nombre descriptivo es “Census tract 55.7”.
- Geometría o polígono que describe los límites de la geografía a nivel geo espacial.
- Tipo de geografía. Esta es una variable añadida manualmente al procesar los datos, para facilitar consultas, identifica si la geografía es una zona censal, condado, etc.

3.2.3. Entidades derivadas de indicadores

Dados los datos de indicadores recolectados, se considera prudente realizar la separación de valores de los indicadores, con la información identificadora de estos, esto lleva a las siguientes entidades:

- **Indicadores:** Entidad que contiene toda la información relevante para identificar los indicadores disponibles en un nivel administrativo. Atributos de esta tabla incluyen la descripción del indicador, un identificador único para este, tipo de unidad de medida, si el indicador es categórico o numérico, tipo de dato del indicador, etc.
- **Indicadores Numéricos:** Tabla que contiene el valor de los indicadores numéricos, incluyendo el identificador geográfico al que corresponde el indicador, y el identificador del indicador al que corresponde el valor.
- **Indicadores Categóricos:** Tabla que contiene los valores de los indicadores categóricos, además de incluir los mismos atributos que los indicadores numéricos, incluye un campo extra para determinar el valor ordinal del valor categórico, si corresponde.

Es de notar la distinción entre indicadores numéricos y categóricos. Si bien se podrían guardar todos los valores en una sola tabla, en este caso se considera mejor mantener cierta consistencia en los datos, de forma que se puedan evitar columnas con valores nulos en múltiples entradas, así como una búsqueda más expedita en el caso de saber que tipo de indicador se busca.

Una decisión de diseño que es importante señalar es que el FEND se almacena como un indicador más en la tabla de indicadores, y los valores de este para cada área se almacenan en la tabla de indicadores numéricos. Esta decisión se debe a que el FEND es, al igual que los indicadores de salud, un indicador a nivel de ambiente, por lo que permite mantener consistencia en la base de datos.

3.2.4. Modelo de datos final

Relacionando las entidades identificadas anteriormente se obtiene el modelo de datos descrito en la figura 3.1. Si bien la mayoría de relaciones son directas y han sido explicadas anteriormente, destaca la relación implícita entre las instalaciones y las geografías, en particular esta relación es espacial, dado que las geografías pueden incluir varios niveles administrativos, una instalación puede pertenecer a varias geografías (correspondientes a distintos tipos de zona administrativa), y, a su vez, una geografía puede tener varias instalaciones.

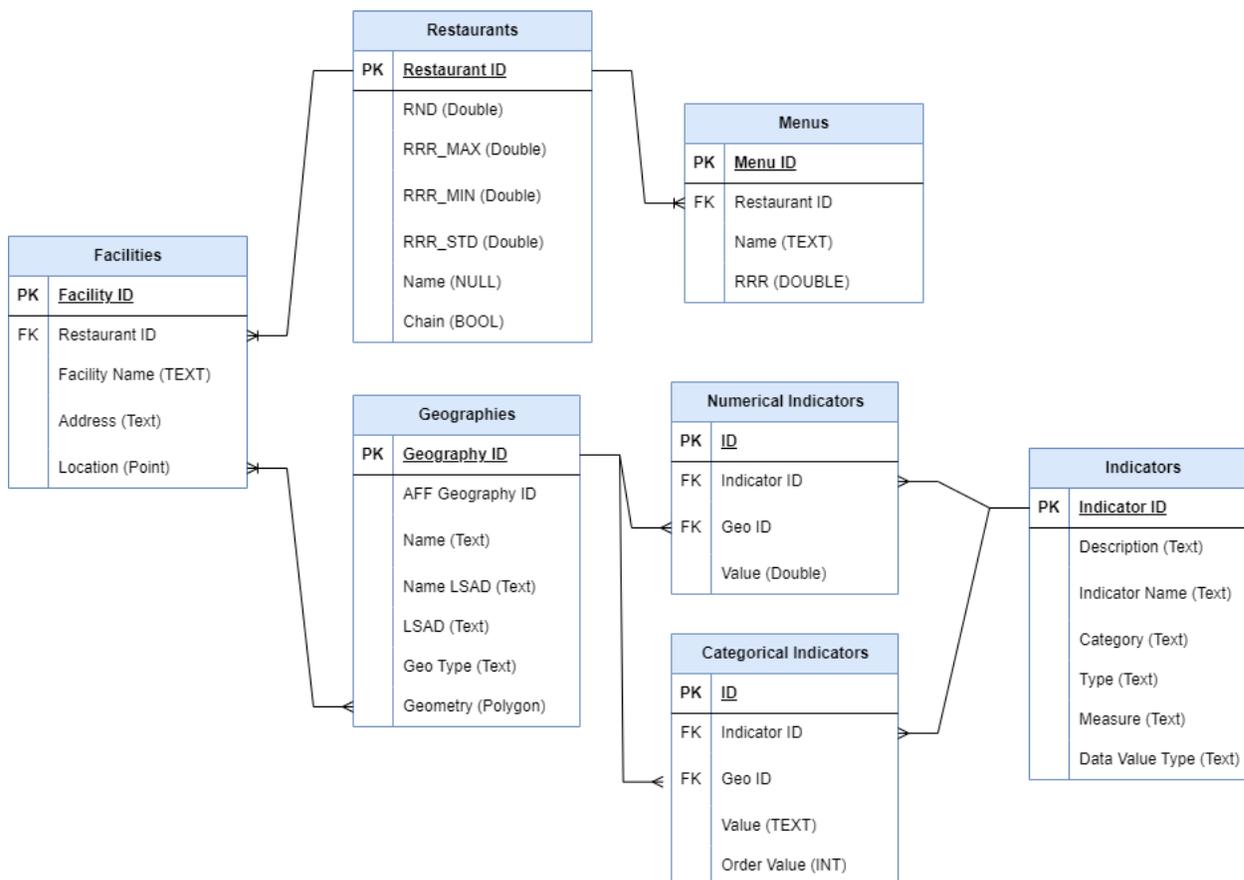


Figura 3.1: Modelo de datos

3.3. Diseño del portal web

Una vez se cuenta con una comprensión adecuada de los datos a utilizar, así como un modelo de datos adecuado, el siguiente paso es diseñar el portal web. El desarrollo del prototipo sigue la metodología de “Diseño Centrado en el Usuario”, también conocido como “User-Centered Design” (UCD) [21]. Esta metodología de trabajo es un proceso iterativo en el que primero se comprende el contexto de uso del software, posteriormente se definen los requerimientos de este, se diseña e implementa la solución, y posteriormente es evaluada por los usuarios. Si en la evaluación de los usuarios, se determina alguna falencia o margen de mejora, se retorna a alguna de las fases anteriores y se mejora la solución desde dicho punto [11], esto es importante en fases posteriores del desarrollo, pues en base al feedback recibido se realizan modificaciones al diseño de la aplicación.

3.3.1. Requisitos de usuario

Los requisitos de usuario se definen directamente de un documento proporcionado por el equipo de la USC, en el cual se detalla el plan y progreso de la investigación para el algoritmo de densidad nutricional, así como la visión que se tiene para la aplicación, siendo esta dirigida principalmente a investigadores y gente involucrada en la nutrición y políticas públicas, aunque con la posibilidad de ser utilizada por cualquiera con acceso a internet. En particular, se especifica directamente que la aplicación debe poseer un mapa interactivo en el que se pueda visualizar la distribución nutricional de los indicadores, siendo este el punto

principal de la aplicación, así como permitir visualizar la relación del indicador nutricional con otras variables de interés, además de la descarga de datos sobre el indicador.

Tomando en cuenta lo anterior, se definen los siguientes requisitos de usuario:

- El usuario debe ser capaz de visualizar la distribución nutricional para un restaurante o área administrativa de su elección.
- El usuario debe ser capaz de visualizar relaciones entre la métrica nutricional de una zona o establecimiento, con las variables de interés definidas.
- El usuario debe ser capaz de comparar las métricas nutricionales entre zonas censales o restaurantes.
- El usuario debe ser capaz de descargar los datos del indicador nutricional al nivel administrativo que sea de su interés, de forma que pueda utilizarlos para sus fines personales.

Con base en estos requisitos, y siguiendo la arquitectura clásica de una aplicación web, se deben escoger tecnologías adecuadas tanto para el backend como el frontend de la aplicación, de forma que se pueda facilitar el desarrollo de ambas partes en lo posible, tomando en cuenta aspectos como el soporte, compatibilidad y librerías de las tecnologías a utilizar.

3.3.2. Arquitectura física

En cuanto a la arquitectura física, se toma en cuenta que la aplicación debe estar disponible en la web y que todos los usuarios deben ser capaces de acceder a los mismos datos y visualizaciones, por lo más sencillo es que el manejo de los datos sea manejado por un único servidor, de forma que la información sea consistente entre todos los usuarios. Debido a esto, se decide que una arquitectura estándar cliente-servidor (figura 3.2) debería ser más que suficiente para el desarrollo de la aplicación, pues permitiría que el manejo de los datos y la entrega de estos sea uniforme entre todos los usuarios.

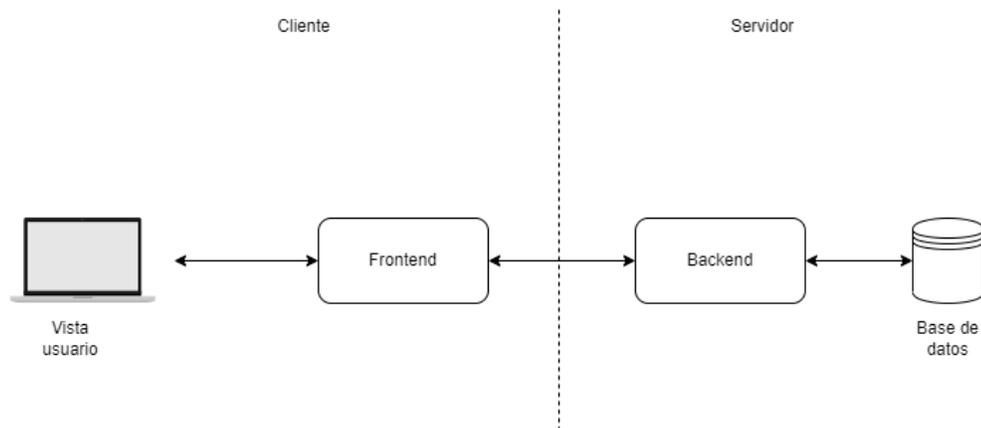


Figura 3.2: Arquitectura simple de cliente-servidor.

3.3.3. Arquitectura lógica

Una arquitectura lógica estándar de 3 capas debería ser más que suficiente para cumplir con las funcionalidades de la aplicación.

- **Capa de presentación:** La interfaz de usuario, a través de la cual se interactúa con la aplicación.
- **Capa de aplicación:** La lógica y procesos que la aplicación lleva a cabo al interactuar con la interfaz.
- **Capa de datos:** La base de datos.

3.3.4. Diseño inicial de la aplicación

3.3.4.1. Navegación y mapa interactivo

La aplicación web debe contar con una interfaz que permita el acceso a cualquiera de las funcionalidades en cualquier momento. Por lo que la inclusión de una barra de navegación sencilla es importante para permitir el acceso a cualquiera de las funcionalidades implementadas.

En el caso del mapa interactivo, este debe contar con zonas coloreadas de acuerdo al valor del indicador en dicha zona, permitir seleccionarlas con el cursor y, en base a dicha selección, mostrar en un *popup* un resumen de los datos en dicha zona, así como la distribución de los indicadores nutricionales definidos. Una idea inicial de visualización básica del mapa, creada con la librería *Kepler.gl* se ilustra en la figura 3.3, en donde se presenta la visualización del indicador a nivel de zona censal, así como la presencia de un *popup* con algunos de los datos disponibles en dicha zona. En este mock-up inicial aún no se diseñan detalles como el controlador del mapa, o las visualizaciones de distribución del indicador por área, esto debido a limitaciones de *Kepler.gl*.

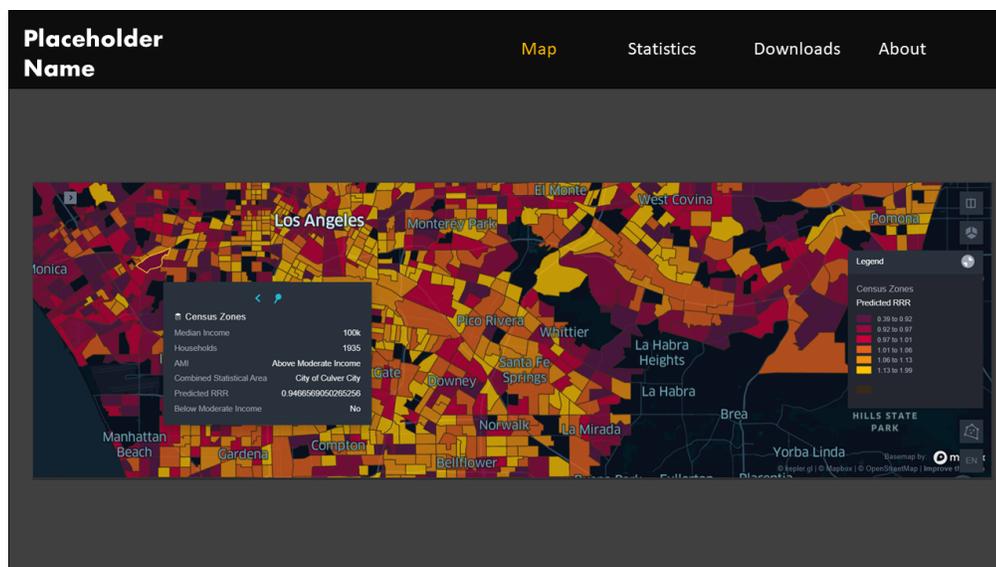


Figura 3.3: Mock-up para vista de mapa interactivo

3.3.4.2. Visualización de estadísticas

En el caso de las estadísticas resumen, estos deben tener su propia sección en la página y obtenerse de acuerdo a ciertos parámetros preestablecidos. Algunas de las visualizaciones a mostrar incluirían la distribución del indicador nutricional en un cierto nivel administrativo,

mostrando su correlación con alguna de las variables de interés. Una idea inicial de visualización de estadísticas se encuentra en la figura 3.4, notar que en el mockup existe una pestaña llamada “RRR distribution”, esta es una idea descartada de mostrar la distribución general del indicador nutricional, dado que se prefirió mostrar esta información por área junto al mapa, para poder realizar la asociación con el espacio físico en el que se calcula el indicador.

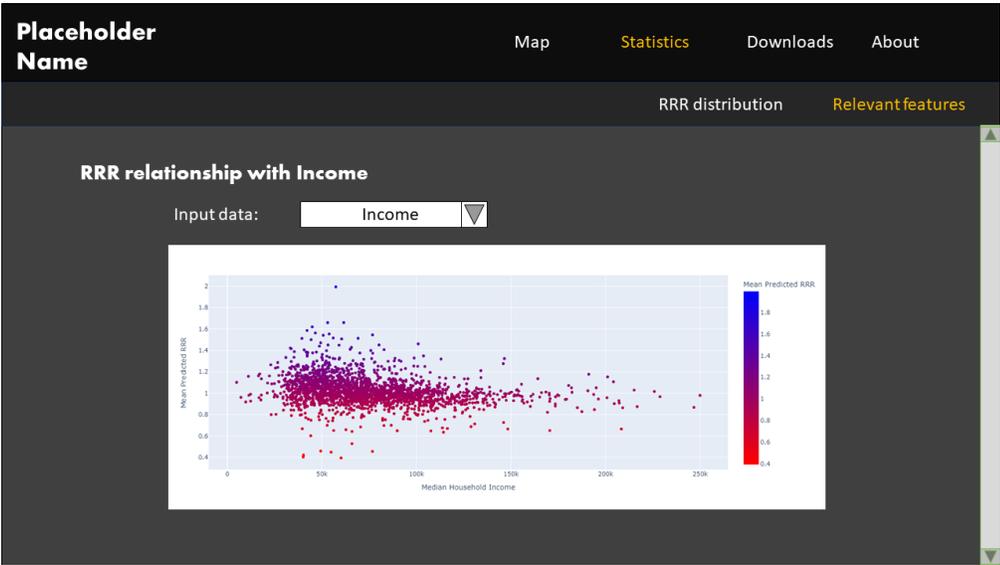


Figura 3.4: Mock-up de idea inicial para vista de visualización de densidad nutricional en base a características de interés.

3.3.4.3. Descarga de datos

Para la descarga de datos se considera habilitar una sección aparte en el portal. Una idea inicial es permitir la descarga de datos a distintos niveles administrativos, permitiendo incluir las variables correlacionadas con los datos del indicador correspondiente, mostrando una previsualización en una tabla. Un mock-up de esta idea se muestra en la figura 3.5.

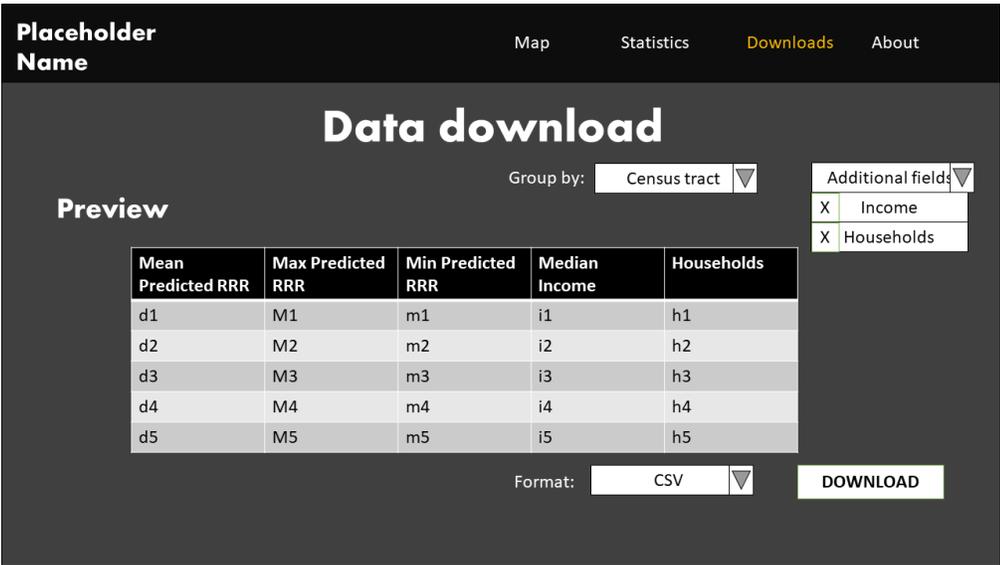


Figura 3.5: Mock-up de idea inicial para vista de descarga de datos.

3.3.4.4. Información adicional

Por último, se estima conveniente que, dado que esta aplicación nace de una investigación, exista una sección del portal en donde se explique brevemente el contexto de la investigación e indicadores. Un ejemplo de lo que se quiere lograr se muestra en la figura 3.6

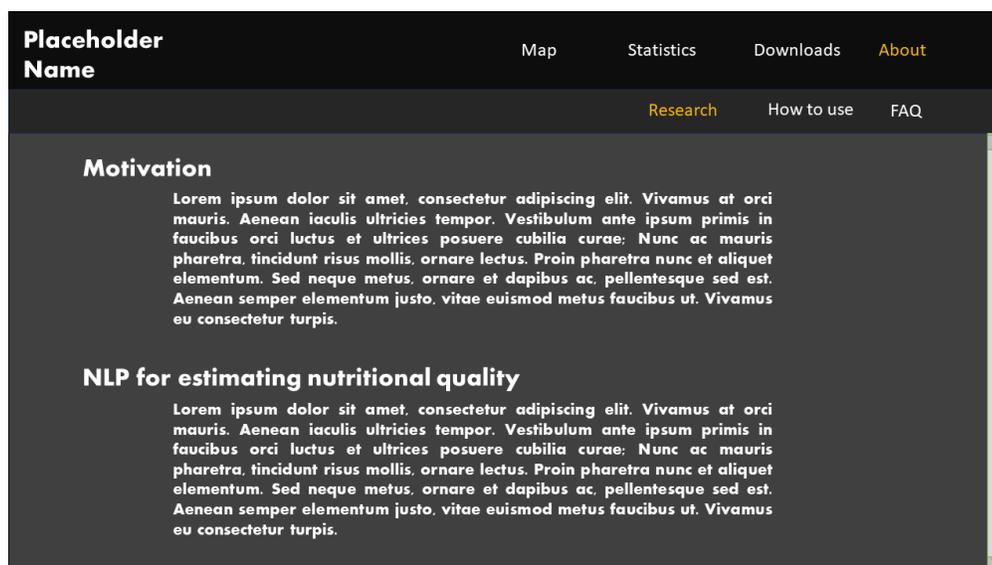


Figura 3.6: Mock-up de ejemplo para sección de información adicional.

3.3.5. Tecnologías

Una parte vital para el desarrollo de la aplicación es la elección de las tecnologías que faciliten y optimicen el desarrollo y uso de la aplicación.

3.3.5.1. Backend y Base de datos

Para la elección de las tecnologías asociadas al tratamiento de datos, el criterio de decisión se basa en las facilidades que dichas tecnologías entreguen para consultar y procesar datos espaciales.

Para la base de datos, se considera ideal utilizar el motor de bases de datos *PostgreSQL*, principalmente debido a su extensión *PostGIS*, la cual permitirá almacenar los datos geoespaciales y realizar consultas optimizadas para estos. De igual forma, para el backend se opta por utilizar el framework *Django*, principalmente debido a la existencia de *GeoDjango*, una extensión del *framework* optimizada para trabajar con datos geoespaciales.

3.3.5.2. Frontend y visualizaciones

Para la elección de tecnologías asociadas directamente con la interfaz de usuario se prioriza la elección de librerías y *frameworks* versátiles y amigables con el usuario.

En este caso, para el *frontend* se considera que *React JS* es una opción viable, debido al amplio soporte, facilidad de desarrollo y compatibilidad con librerías de visualización. En el caso de los mapas interactivos, se optó por utilizar la librería *Deck.gl*, principalmente por su directo uso con *React JS*, la facilidad para cargar datos y la posibilidad añadir eventos

personalizados al interactuar con el mapa. Por último, para la visualización de estadísticas se utiliza la librería *Plotly*, dado que permite crear gráficos de distribución de forma sencilla.

Capítulo 4

Implementación de la solución

4.1. Base de datos

La implementación de la base de datos implica, además de los procesos típicos de construcción de *schema* y tablas, el preprocesamiento de los datos crudos para adaptarlos al modelo de datos diseñado. A continuación se describe a grandes rasgos el preprocesamiento realizado, así como la población de la base de datos.

4.1.1. Preprocesamiento de datos

Para el preprocesamiento de datos se utiliza el lenguaje *Python*, utilizando las bibliotecas *Pandas* y *GeoPandas*.

En el caso de datos que no requieren mucho preprocesamiento, debido a que presentan una estructura similar al modelo de datos, el preprocesamiento se limita a seleccionar las columnas relevantes, renombrarlas en caso de ser necesario y filtrar los datos de acuerdo a lo requerido por la aplicación. Por ejemplo, en el caso de los datos geográficos, se seleccionan las columnas requeridas por el modelo, se seleccionan solamente los datos correspondientes al estado de California y posteriormente se concatenan los *dataframes* de cada nivel administrativo (pues los datos de cada nivel se descargan por separado).

En el caso de datos que requieran de normalización, el proceso consiste en extraer las columnas y/o datos requeridas para cada tabla y posteriormente eliminar posibles entradas duplicadas. Por ejemplo, en el caso de los datos del indicador de densidad nutricional provenientes de Brizo, es necesario separar los datos de menús de los datos del restaurante y, a su vez, separar los datos de longitud y latitud de los datos del restaurante. Al realizar esta separación, la tabla de restaurantes queda con varias entradas duplicadas que posteriormente son eliminadas, además, tanto los menús como las ubicaciones conservan la columna de ID del restaurante, permitiendo recrear los datos originales en caso de ser necesario.

4.1.2. Población de base de datos

La subida de datos se realiza mediante *Pandas* y *GeoPandas*, utilizando los métodos “*to_sql*” y “*to_postgis*” respectivamente, los cuales permiten la subida directa de datos de *dataframes* a una base de datos, siempre y cuando exista una conexión establecida mediante la biblioteca *SQLAlchemy*. Es importante notar que en el caso de tablas que poseen llaves

foráneas, primero se consultan las tablas respectivas directamente de la base de datos, utilizando la lectura directa desde la base de datos que incluye la biblioteca *Pandas*, luego se asegura que la tabla a ingresar contenga solamente entradas cuyas llaves foráneas ya existen en la base de datos, y posteriormente se ingresan.

Una vez subidos los datos, es de notar que en el preprocesamiento inicial no se han generado los indicadores agregados (RND, FEND, etc), estos se crean mediante consultas SQL en la base de datos. La razón de esto se debe a que, en el caso de añadir más datos para un restaurante o área administrativa, el cálculo podrá tomar en cuenta los datos ya existentes.

4.2. Backend

En el caso del backend, se habilitan endpoints para cada consultar los datos de cada modelo que la aplicación requiera mediante *requests* de tipo GET. En algunos casos se habilita el uso de filtros en base a parámetros de consulta, con el objetivo de agilizar consultas a datos específicos, por ejemplo, al acceder al endpoint `/api/menus/?establishment_id=X`, se obtiene solamente los menús para el establecimiento con ID=X.

El único detalle relevante de implementación se da en la serialización de los datos, pues si bien Django posee funciones para llegar a cabo la conversión de los datos a los formatos CSV y GeoJSON, la serialización a este segundo formato puede llegar a ser bastante lenta, llegando a tardar hasta 15 segundos en serializar 40000 restaurantes en un entorno de desarrollo local. El formato GeoJSON es utilizado para crear directamente las visualizaciones en el mapa, por lo que es indispensable que los datos se obtengan en el menor tiempo posible, es debido a esto que se debió recurrir a utilizar anotaciones de Django para transformar, a nivel de base de datos, las geometrías de los datos geográficos en un polígono reconocible para el formato GeoJSON, y posteriormente llevar a cabo la serialización de los datos a la estructura GeoJSON en el frontend. Al realizar esta modificación al proceso estándar de Django se reducen drásticamente los tiempos de obtención de datos, debido a que se reduce el procesamiento de datos realizado en *Python*, tardando menos de 500ms entre consulta y modificación de datos en el backend, en un entorno de desarrollo local.

4.3. Interfaces

A continuación se describen las vistas implementadas, se explican brevemente algunas de las decisiones de diseño tomadas al implementarlas, así como la interacción esperada de un usuario con estas.

4.3.1. Mapa interactivo

El mapa interactivo es la vista por defecto del portal (figura 4.1), en esta se puede acceder a cualquiera de las otras secciones de la aplicación, o interactuar con el mapa para revisar la distribución del indicador.

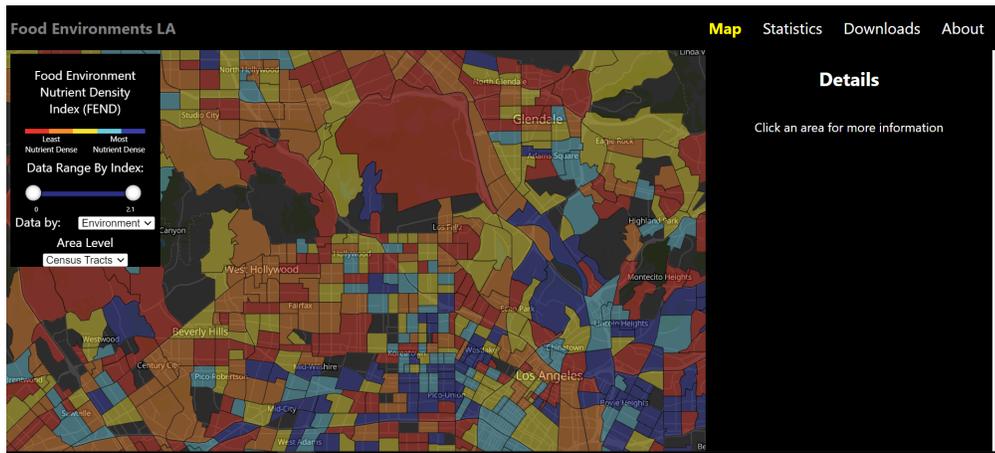


Figura 4.1: Vista inicial del mapa interactivo

El mapa interactivo cuenta a su vez con 3 componentes principales:

1. **Mapa:** La componente principal de la vista, cuenta con zonas coloreadas en base al valor del indicador nutricional en dicha zona, siendo el coloreo por quintiles.
2. **Controlador:** Componente posicionada en la parte superior izquierda del mapa. Permite modificar el mapa según el indicador requerido (RND o FEND), así como los datos que se muestran.
3. **Barra de detalles:** Componente posicionada al lado derecho del mapa, muestra detalles estadísticos sobre la zona seleccionada en el mapa.

En esta vista, el flujo de usuario esperado parte de mover el cursor sobre las zonas coloreadas del mapa, tras lo cual aparece un *popup* con un identificador de la zona seleccionada, así como las estadísticas del indicador en dicha zona (figura 4.2). Si bien se consideró en un momento incluir estadísticas de las variables correlacionadas en dicho *popup*, se consideró que sería mejor mostrar dichos datos exclusivamente en la sección de estadísticas, de forma que el mapa se enfoque exclusivamente en presentar información de densidad nutricional. Una vez

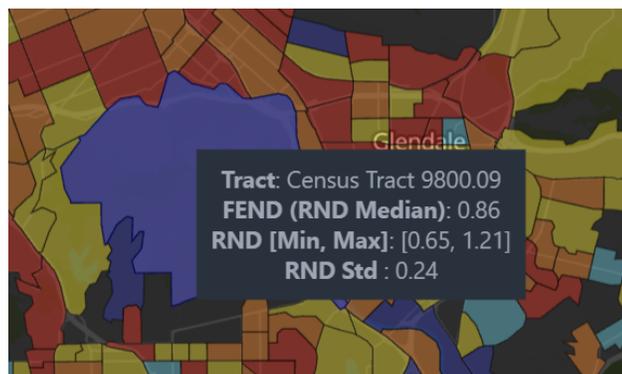


Figura 4.2: Ejemplo de popup en mapa interactivo

el usuario encuentra una zona de su interés, puede clicar en esta para desplegar los detalles de distribución del indicador correspondiente en la barra lateral, así como un resumen de los menús encontrados en la zona seleccionada, en forma de una nube de palabras (figura 4.3).

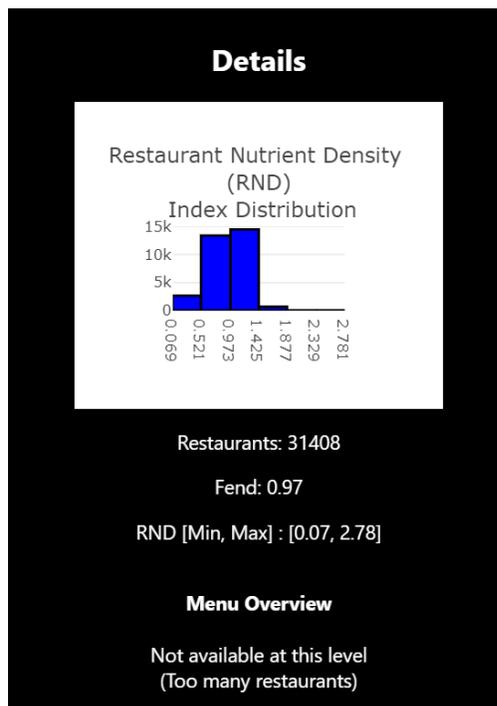


Figura 4.4: Ejemplo de detalles en barra lateral, cuando no es posible procesar la cantidad de menús

Como última componente del flujo de usuarios, este, mediante el controlador del mapa (figura 4.5), puede decidir el tipo de unidad (restaurantes o ambientes) a visualizar en el mapa, así como filtrar las unidades mostradas dependiendo del valor del indicador.

- Si el usuario decide visualizar por restaurantes, se presentarán los restaurantes coloreados por sus respectivos RND. El usuario puede decidir si quiere visualizar todos los restaurantes, o solamente los que pertenecen (o no) a una cadena de restaurantes. En la figura 4.6 se puede apreciar la vista por restaurantes, mostrando solo los restaurantes de cadena.
- Si el usuario decide visualizar por entorno, se presentarán las áreas administrativas coloreadas por sus respectivos FEND. El usuario puede decidir el tipo de área administrativa a mostrar. En la figura 4.7 se puede apreciar la vista por entorno, utilizando el nivel administrativo *places*.

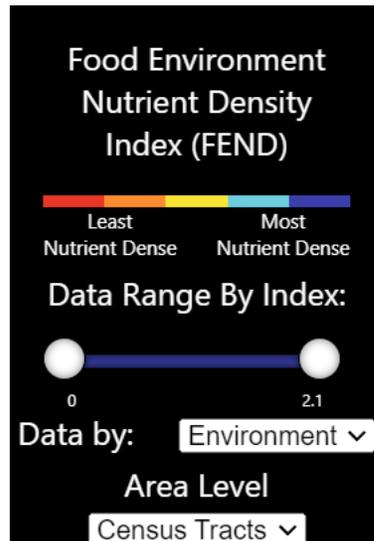


Figura 4.5: Controlador del mapa

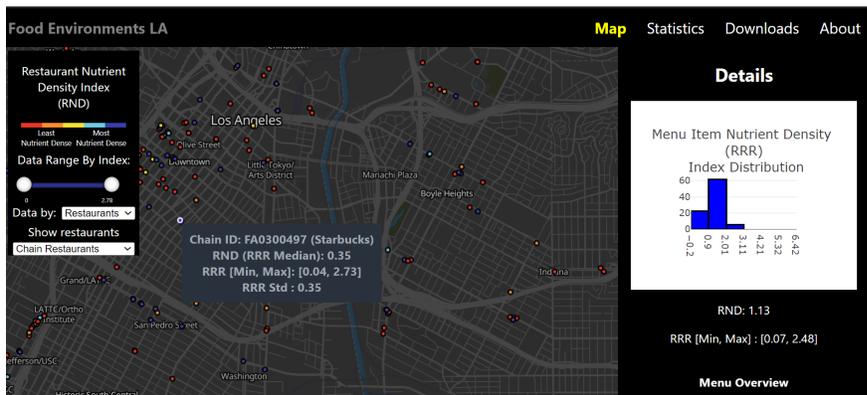


Figura 4.6: Mapa interactivo con indicador por restaurantes, mostrando solamente los restaurantes de cadena.

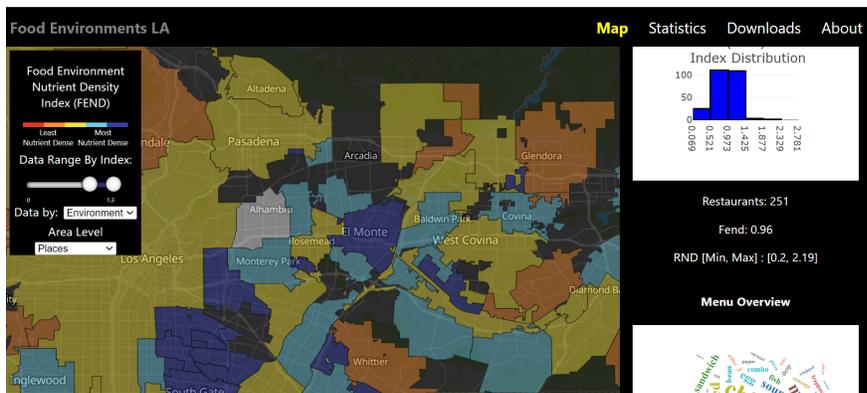


Figura 4.7: Mapa interactivo con indicador por ambientes, a nivel *places*, filtrando las zonas de baja densidad nutricional.

4.3.2. Estadísticas

En esta vista, el usuario es capaz de visualizar como se correlacionan los indicadores nutricionales de ambiente con las variables de interés recolectadas. Por defecto, el usuario visualiza las variables a nivel de zona censal, pudiendo escoger cualquiera de las variables de interés (figura 4.8).

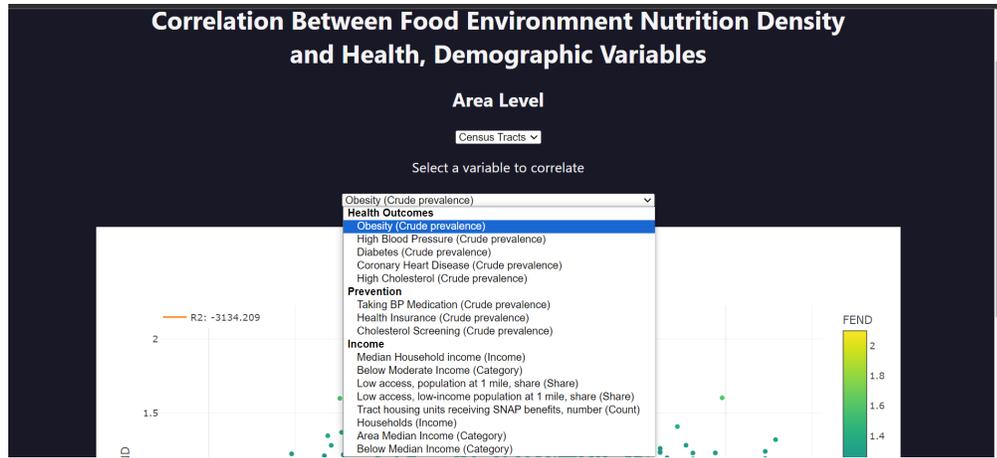


Figura 4.8: Vista por defecto de las estadísticas, junto a la selección de variables.

Cuando se trata de indicadores numéricos, el usuario visualiza un *scatterplot* de la distribución de la variable, siendo el eje *X* el valor de la variable de interés, y el eje *Y* el valor del FEND. Para poder analizar mejor como se correlacionan los datos, esta visualización incluye la línea resultante de aplicar una regresión lineal con los datos, junto al respectivo valor del coeficiente R^2 . Un ejemplo se muestra en la figura 4.9.



Figura 4.9: Correlación entre el FEND y la prevalencia de presión alta en adultos, a nivel de lugares (*places*).

En el caso de las variables categóricas, el usuario puede alternar entre un *barplot* (figura 4.10) con la mediana del FEND en cada categoría, y un *boxplot* que muestra la distribución del FEND en cada categoría (figura 4.11).



Figura 4.10: Correlación entre el FEND y la prevalencia de presión alta en adultos, a nivel de lugares (*places*).

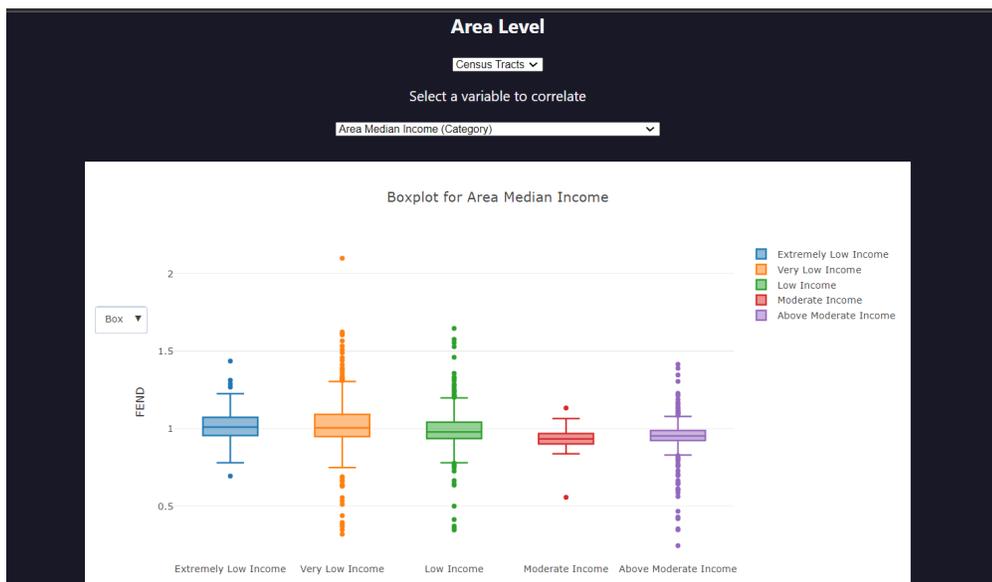


Figura 4.11: Correlación entre el FEND y la prevalencia de presión alta en adultos, a nivel de lugares (*places*).

4.3.3. Descargas

En esta vista se habilita la descarga del indicador nutricional RND y FEND en todos sus niveles disponibles, permitiendo la descarga en formatos CSV (incluyendo el identificador geográfico o latitud y longitud) y GeoJSON (incluyendo la geometría correspondiente).

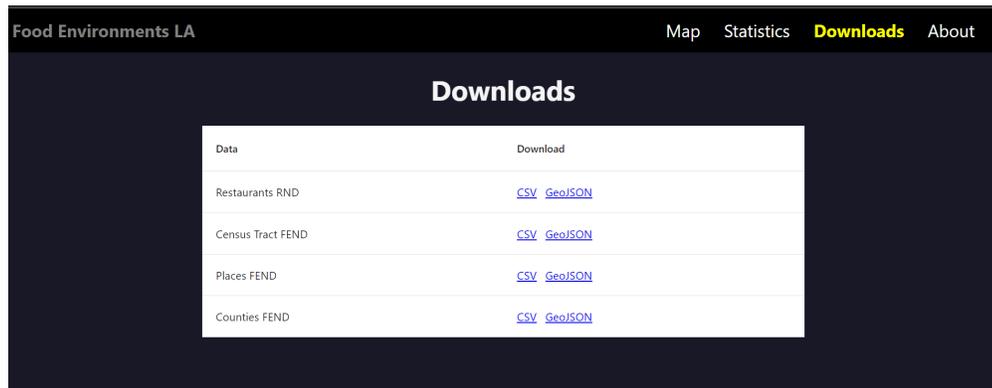


Figura 4.12: Vista de descarga de datos

Si bien en el diseño inicial se contemplaba la posibilidad de juntar los datos del indicador nutricional con las variables de interés utilizadas en las estadísticas, finalmente por falta de tiempo, para disminuir el peso de los archivos, y debido a que los otros indicadores son datos públicos, se optó por simplificar la descarga de datos, incluyendo solo el indicador nutricional adecuado junto a los datos necesarios para identificar la ubicación a la que corresponde.

4.3.4. Información

Esta sección simplemente consiste en una sola página, con una descripción de los indicadores y la investigación realizada por el equipo de la Universidad del Sur de California, así como reconocimientos a las organizaciones involucradas y miembros del equipo (figura 4.13).

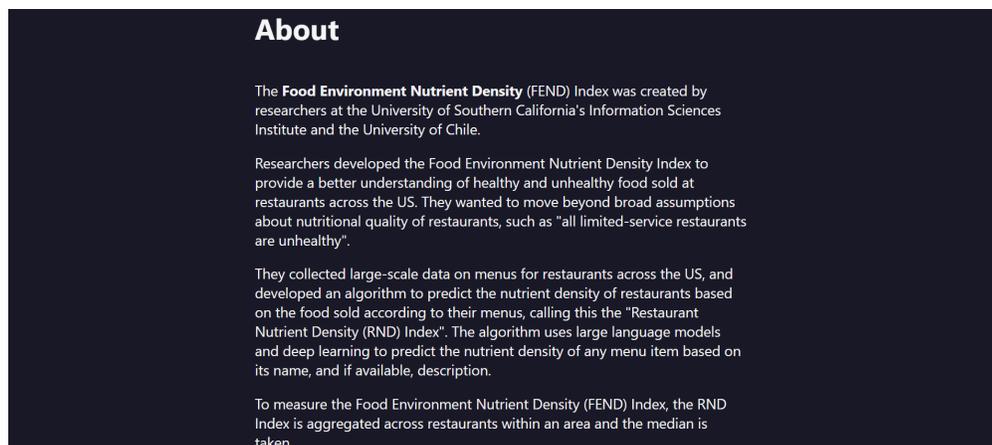


Figura 4.13: Vista de sección de información

Capítulo 5

Validación

A continuación se expone la opinión de usuario sobre la plataforma, esta opinión está basada en el despliegue de la aplicación en uno de los servidores del DCC [17]. La evaluación fue realizada por el equipo de la USC, siendo la evaluadora la investigadora Abigail Horn. Se considera oportuno mencionar que existieron dos instancias principales en las que se comentó el funcionamiento de la plataforma, en la primera se dio retroalimentación sobre posibles cambios en una versión preliminar de la plataforma, y en la segunda se respondieron preguntas sobre usabilidad y eficiencia, y se dieron comentarios sobre la versión más actualizada de la aplicación, la cual es la utilizada de ejemplo durante la sección de implementación de este informe.

5.1. Pruebas de validación

Antes de hablar sobre la validación de la aplicación en sus dos iteraciones, se describirá brevemente el tipo de tareas ejecutadas en la aplicación, así como las preguntas respondidas en la segunda iteración.

Si bien las pruebas y validación de la aplicación fueron realizadas por el principal grupo objetivo de usuarios, el equipo de investigación de la USC, con Abigail Horn como evaluadora, se cree que en futuras iteraciones, cuando se disponga a desplegar oficialmente la aplicación en la web, sería prudente realizar una validación más rigurosa con un conjunto más amplio y diverso de usuarios, desde gente asociada con la investigación, hasta usuarios que no posean ningún conocimiento previo sobre esta. En este sentido, las tareas y preguntas expuestas no sólo deberían considerarse como un registro y ejemplo de lo realizado, sino como la base de una planificación para futuras pruebas de validación

5.1.1. Tareas ejecutadas

En el proceso de validación, se probaron las funcionalidades básicas de la aplicación. En esta sección se exponen las tareas básicas ejecutadas en las pruebas, las cuales si bien no necesariamente fueron ejecutadas en el orden expuesto, se exponen en orden como un posible ejemplo a seguir en el caso de futuras pruebas de validación, basadas en lo que se espera de un usuario promedio.

5.1.1.1. Tareas en sección inicial

1. Ingresar a la aplicación.

2. Seleccionar una zona y visualizar las estadísticas resumen.
3. Escoger una visualización distinta a nivel de área y repetir el paso 2.
4. Manipular el slider de la aplicación y visualizar el efecto del filtro en el mapa.
5. Cambiar el nivel de visualización de los datos, de área a restaurante.
6. Repetir pasos 2 y 4 a nivel de restaurante.

5.1.1.2. Tareas en sección de estadísticas

1. Seleccionar sección de estadísticas.
2. Seleccionar una variable a relacionar, observar la visualización correspondiente.
3. Cambiar el nivel de área de visualización.
4. Repetir pasos 2 y 3 las veces que se estime conveniente.

5.1.1.3. Tareas en sección de descargas

1. Seleccionar sección de descargas.
2. Seleccionar un archivo a elección para descarga, en formato CSV, y esperar a que la descarga se complete.
3. Seleccionar un archivo a elección para descarga, en formato GeoJSON, y esperar a que la descarga complete.
4. Repetir pasos 2 y 3 con archivos distintos, las veces que se estime conveniente.

5.1.1.4. Tareas en sección de información

1. Seleccionar sección de información.
2. Leer texto presente.
3. Acceder a los enlaces que se presentan.
4. Repetir paso 3, las veces que se estime conveniente

5.1.2. Preguntas de validación

Una vez un usuario ha ejecutado las tareas, se debe consultar sobre el correcto funcionamiento de las funcionalidades expuestas en las tareas ejecutadas, a su vez, se deben realizar una serie de preguntas para asegurar la utilidad y usabilidad de la aplicación. En el contexto de la validación realizada, algunas de estas preguntas fueron realizadas durante la segunda iteración de la aplicación, mientras que otras son preguntas que no se consideraron en su momento, pero podrían ser útiles en futuras pruebas.

5.1.2.1. Preguntas enfocadas a investigadores

Las siguientes preguntas de usabilidad van enfocadas principalmente a gente con conocimiento previo de la investigación realizada, y fueron respondidas por la investigadora Abigail Horn durante el proceso de validación. Estas preguntas tienen el objetivo de identificar el potencial de la aplicación para dar a conocer la eficiencia y caracterización del algoritmo de densidad nutricional, así como para buscar retroalimentación general de la aplicación.

1. ¿La información que provee el mapa ayuda en la identificación de entornos o restaurantes con alta o baja densidad nutricional?
2. ¿Las visualizaciones estadísticas ayudan a corroborar la existencia o no existencia de relaciones entre el indicador FEND y las variables de interés?
3. ¿Los datos disponibles para descarga poseen la información necesaria para que un usuario externo a la investigación pueda trabajar con los indicadores?
4. ¿Los datos se cargan en un tiempo prudente?
5. ¿Que sería bueno añadir, remover o mejorar para futuras iteraciones del proyecto?

5.1.2.2. Preguntas para usuarios ajenos a la investigación

Dado que el enfoque principal del trabajo de título fue desarrollar un prototipo adecuado para los estándares del equipo de investigación, sólo se tomó en cuenta a este conjunto de usuarios para la validación. Sin embargo, en retrospectiva se considera que se dejó de lado la opinión de posibles usuarios ajenos a la investigación realizada, que si bien no se encuentran en el grupo principal de potenciales usuarios (investigadores y gente con conocimientos del área), sí podría ser un conjunto de usuarios valioso para brindar una nueva perspectiva sobre el proyecto.

En este contexto, se considera que en futuras pruebas de validación se podrían realizar las siguientes preguntas enfocadas en este tipo de usuario, con un enfoque en la presentación de la aplicación y facilidad de uso para alguien sin experiencia en el tema.

1. ¿La aplicación provee información suficiente para aprender y comprender la investigación realizada y los indicadores desarrollados?
2. ¿Consideras que las visualizaciones y datos presentados son comprensibles para alguien que acaba de aprender sobre los indicadores desarrollados y su contexto?
3. ¿Consideras que el estilo visual de la aplicación permite navegar intuitivamente por esta, y visualizar la información fácilmente?

5.2. Primera iteración

Durante la primera iteración se remarcó la lentitud de la aplicación al cargar los datos del mapa, se recalca que en este primer despliegue se estaba utilizando la serialización por defecto de Django, esto sumado a que la carga de datos suele ser más lenta en un servidor, dió como resultado que los datos de restaurantes tardaran hasta varios minutos en cargar. Este problema se tomó en cuenta para agilizar la entrega de los datos en el backend, como

se mencionó en la sección de implementación.

También se recalcaron detalles de interfaz de usuario, varios con detalles menores como cambiar la leyenda de ciertos datos. A continuación se listan los comentarios más relevantes:

- “Please make RND distribution x-scale across restaurants”. Este comentario se refiere a que, en esta primera iteración, las escalas de distribución del indicador para distintos restaurantes y/o áreas administrativas solo tomaban en cuenta su distribución local. Así que por ejemplo, un restaurante podía presentar una escala de distribución de entre 1 y 1.5 para sus menús, mientras que otro podía presentar una escala entre 0.8 y 1.2. Esto se arregló simplemente tomando en cuenta el mínimo y máximo global de los indicadores correspondientes (RND en el caso de áreas administrativas, RRR en el caso de restaurantes).
- “Create a toggle to select that allows visualizing, separately, the Brizo data (all restaurants, no names), and Spoonacular (only chains, having names) data”. Este comentario se refiere a que en esta primera iteración, al seleccionar los datos de restaurantes simplemente se mostraban todos los datos disponibles. Como se mencionó en la implementación, esta funcionalidad fue añadida, de forma que se permite visualizar todos los datos, solo aquellos en los que los restaurantes poseen nombre, y solo aquellos en los que los restaurantes pertenecen a una cadena.
- “Please allow selection by Census Tract or by Neighborhood (this area level data will be shared)”. Este comentario se refiere a que en esta primera iteración, a nivel de área solo se mostraban los datos de zonas censales. Es debido a esto que se añadieron otros niveles administrativos durante la implementación, sin embargo, al no existir datos oficiales de vecindarios en *Census.gov*, y al no ser compartidos estos datos a fecha de escritura de este informe, que simplemente se añadieron los niveles administrativos que se consideraran prudentes, para asegurar que la aplicación sea capaz de manejar datos de múltiples niveles administrativos en un futuro.

A pesar de las críticas se recalcaron aspectos positivos, como el aspecto visual del mapa, la existencia de la barra lateral y los datos que presenta, o la integración de las variables de interés para la sección de estadísticas.

5.3. Segunda iteración

Para la segunda iteración de la aplicación, se realizaron una serie de preguntas a Abigaíl Horn, investigadora de la USC, para corroborar la usabilidad, utilidad y eficiencia del prototipo. La transcripción textual de las preguntas realizadas, junto a sus respectivas respuestas, se encuentra en el anexo A.

5.3.1. Mapa interactivo

Se consultó sobre la utilidad de la información provista por el mapa, en particular para detectar entornos y/o restaurantes de baja o alta densidad nutricional. La respuesta fue favorable, aludiendo a que la visualización ayuda a descubrir áreas con un bajo acceso a nutrientes que no sería posible identificar con métodos tradicionales. Citando textualmente: “*Yes, the information in the visualization helps these users to identify the locations of higher*

vs. lower nutrient density restaurants and food environments, helping to uncover areas of particularly low nutrient access that are not possible to identify with traditional category-based definitions of food environment nutritional quality.”

5.3.2. Visualizaciones estadísticas

Se consultó sobre la utilidad de las visualizaciones estadísticas implementadas en la aplicación, en particular aquellas que relacionan el indicador FEND con las variables de interés. La respuesta fue favorable, aludiendo al hecho de que las visualizaciones son lo suficientemente claras para dar una intuición inicial sobre la correlación entre el FEND y las variables de interés, en particular cuando se tenga una versión más actualizada del primero. Citando textualmente: *“It is currently hard to say whether the bivariate statistical results help to corroborate the existence/non-existence of relationships between the FEND indicator and the variables of interest, since the FEND data is not up to date. However, the statistical results are very clearly visualized so as soon as the data are updated it should be possible to assess correlations between variables of interest. More comprehensive multivariable statistical models are ultimately needed to determine associations between the FEND indicator and variables of interest, but the bivariate correlations can provide an initial intuition for the direction of results.”*

5.3.3. Carga y descarga de datos

Debido a los problemas de carga de datos presentados en la primera iteración, se consultó sobre la eficiencia actual de carga de estos. La respuesta fue favorable, sin menciones de nuevos problemas en la carga de datos. Citando textualmente *“The data are delivered in timely manner.”*

También se consultó sobre la utilidad de los datos entregados en la sección de descargas, en particular si se consideraba que los campos entregados eran los suficientes y apropiados para que un usuario cualquiera pudiera trabajar con estos. La respuesta fue favorable, aunque recalcando detalles a mejorar en los archivos descargados. En particular, se mencionaron detalles como el nombre y orden de los atributos, los cuales podían ser poco intuitivos para el usuario, también surgieron sugerencias como redondear los indicadores para no dar una falsa sensación de precisión, o la adición de un archivo README que explique las variables.

5.3.4. Posibles mejoras

Se consultó sobre posibles mejoras o adiciones a posibles iteraciones futuras del prototipo desarrollado. A continuación se listan estas:

- Se recalcaron detalles como mejoras de usabilidad de cara al usuario, como por ejemplo un tutorial sobre como utilizar y entender las visualizaciones.
- Integración del modelo de densidad nutricional al portal, de forma que un usuario pueda ingresar el nombre de un menú y obtener la composición nutricional estimada, junto al índice RRR.
- Funcionalidades de búsqueda para geografías o restaurantes específicos, o incluso en base a descripciones de menú.

- Realzar casos de estudio en áreas de alta densidad nutricional comparado con baja densidad nutricional.
- Integración de una mayor cantidad de variables de interés.

5.4. Discusión

5.4.1. Cumplimiento de objetivos

Tomando en cuenta los comentarios realizados por Abigail Horn, se puede concluir que la aplicación logra el objetivo inicial de entregar visualizaciones útiles para conocer la distribución y relación de los indicadores desarrollados. Adicionalmente, se lograron corregir los problemas iniciales presentados en la primera iteración, como la lenta carga de los datos o la falta de filtros de selección categóricos en el mapa, los cuáles son aspectos que mejoraron en la segunda iteración y no volvieron a ser mencionados como posibles problemas.

A pesar de que el prototipo logra cumplir funcionalmente los objetivos y requisitos planteados, también queda en evidencia que se podría haber implementado una mejor retroalimentación y guía de cara a usuarios nuevos, así como funcionalidades para facilitar el uso de la herramienta. Sugerencias como las mencionadas para la sección de descargas, dejan en evidencia esto, en donde cambios simples como la adición de un archivo README o el reordenamiento de variables en los archivos descargados podrían haber resultado en una mejora de la experiencia de usuario.

5.4.2. Mejoras de optimización

Se quiere recalcar la importancia de buscar alternativas a los mecanismos por defecto utilizados en las tecnologías escogidas. El ejemplo claro de esto es la serialización de datos realizada por Django, en donde se logró ahorrar una cantidad considerable de tiempo de procesamiento en el backend al cambiar el enfoque por uno no convencional. Se puede asociar el bajo rendimiento del método convencional al hecho de que *Python* suele ser un lenguaje de programación con un tiempo de ejecución mucho mayor a otros, por lo que realizar la serialización, un proceso que consiste en transformar todas las entradas obtenidas desde la base de datos a un formato directamente comprensible por el frontend, en *Python* puede no ser siempre óptimo, en especial cuando se debe procesar una cantidad masiva de datos. Aún así no se quiere desestimar el uso de Django como tecnología de backend, dado que proporciona muchas facilidades, como el modelo ORM, para implementar una API REST funcional de forma sencilla.

5.4.3. Apreciación personal

Si bien en general la evaluación de la aplicación fue positiva, personalmente se cree hay aspectos no mencionados que podrían mejorarse.

En primer lugar, la aplicación fue diseñada pensando exclusivamente en la experiencia de un usuario de computadora, por lo que no se adapta bien al uso en entornos similares a los de un *Smartphone* (figura 5.1). No se desarrolló una solución a este problema, dado que no fue

algo considerado hasta las etapas finales del desarrollo, sin embargo, se cree que para futuras iteraciones de la aplicación sería importante tener este aspecto en cuenta.

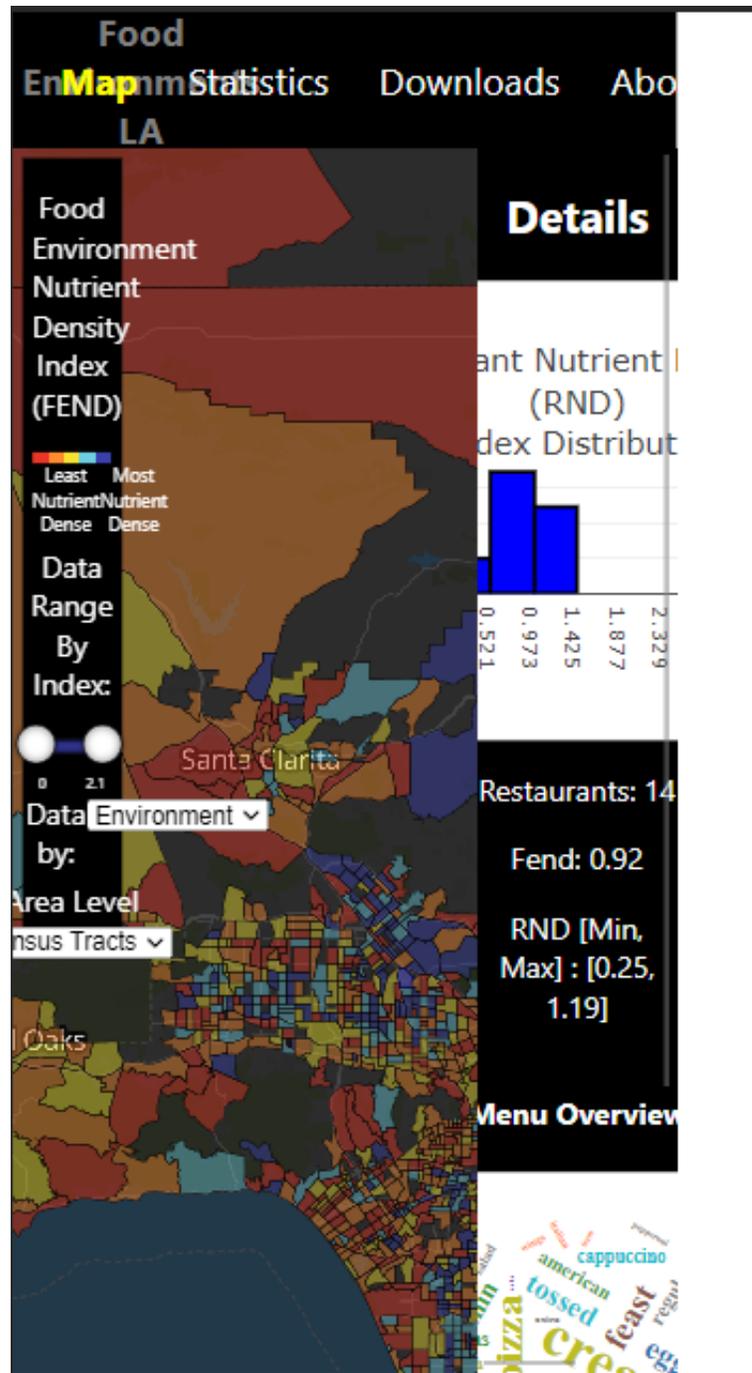


Figura 5.1: Vista por defecto del mapa, en vista simulada de un Samsung Galaxy S8 (Resolución 360 × 740)

Otro aspecto que se considera mejorable es el uso de estilos, en particular la falta de estilos modernos agradables a la vista de un usuario promedio, dado que los estilos utilizados en la aplicación son básicos y se limitan a controlar aspectos como el tamaño de letra y posicionamiento de elementos en la interfaz.

Es de recalcar que todos estos aspectos mencionados podrían mejorar la experiencia de uso de la aplicación, sin embargo, no fueron tomados en cuenta debido a que el enfoque durante el desarrollo priorizó entregar un producto mínimo viable que cumpliera con las funcionalidades requeridas. Proceso en el que se tuvieron que aprender múltiples tecnologías, así como mejorar constantemente las funcionalidades de la aplicación.

Capítulo 6

Conclusión

Para finalizar el presente informe, se concluirá sobre los aspectos más relevantes del trabajo realizado, siendo estos el cumplimiento de los objetivos planteados, así como el trabajo futuro requerido para mejorar lo desarrollado.

6.1. Cumplimiento de objetivos

El objetivo general del trabajo de título presentado era implementar un prototipo capaz de generar conocimiento sobre la calidad y desigualdad nutricional en distintas zonas de la ciudad de Los Angeles y alrededores. Para ello se presentaría información de interés a través de visualizaciones interactivas e indicadores nutricionales. En la sección 1.4.1 se plantearon los objetivos específicos necesarios para lograr el objetivo general, los cuales serán abordados por separado a continuación.

El primer objetivo específico era generar un mapa interactivo que incluyera información útil sobre la distribución de los indicadores en áreas seleccionadas por el usuario. Según los comentarios presentados en la sección de validación, este objetivo fue completado, siendo el mapa implementado y la información proporcionada por este considerada útil para identificar áreas de alta o baja densidad nutricional, las cuales no serían identificables con métodos tradicionales.

El segundo objetivo específico consistía en generar visualizaciones que permitieran conocer la distribución del indicador, así como su relación con las variables de interés. Este objetivo también fue completado, implementando visualizaciones para la distribución del indicador en la barra lateral del mapa interactivo, así como su relación con las variables de interés implementada en la sección de estadísticas del portal. Adicionalmente, según los comentarios presentados en la validación, las visualizaciones estadísticas fueron consideradas útiles para poder generar intuición sobre las relaciones entre el indicador FEND y las variables de interés.

El tercer objetivo específico consistía en permitir la descarga de datos sobre el indicador. Este objetivo fue completado funcionalmente, permitiendo la descarga exitosa de los datos en múltiples niveles administrativos y en dos formatos distintos, CSV y GeoJSON. Aún así, según los comentarios presentados en la sección de validación, se podría haber mejorado esta función para entregar los datos de forma más ordenada, así como guiar al usuario en el significado de las variables.

En general, los objetivos planteados se completaron funcionalmente de forma exitosa. Aunque se pudieron haber realizado modificaciones para que el uso de estas características fueran más amigables e intuitivas para el usuario.

6.2. Trabajo futuro

Tanto el frontend como el backend del proyecto, junto a sus códigos asociados, se dejarán disponibles en un repositorio para el equipo de la USC, de forma que el prototipo pueda ser mejorado en el futuro. Si bien la recepción del prototipo desarrollado fue en su mayoría favorable, y se considera que este es una buena base sobre la cual ir construyendo una aplicación más profesional y robusta. Es importante detallar cuáles son los posibles aspectos a mejorar para que la calidad del producto sea superior en el futuro.

En cuanto a futuras iteraciones del proyecto, se considera prioritario primero abordar cambios estéticos que mejoren la experiencia de usuario, posteriormente la adición de cambios funcionales que faciliten la usabilidad de la aplicación, y por último la expansión del portal con funcionalidades innovadoras.

En cuanto a cambios estéticos, los primeros pasos incluirían mejoras en la retroalimentación y guía de un usuario normal en el sitio web. Estos cambios incluirían los detalles mencionados previamente sobre la sección de descargas, así como el uso de estilos más agradables a la vista y la adaptación del portal a pantallas de *smartphone*.

En cuanto a funcionalidades enfocadas a la usabilidad, se considera que una funcionalidad de búsqueda de geografías basada en el centroide de esta, o la búsqueda de un restaurante específico basado en las coordenadas o nombre de este, permitirían que los usuarios pudieran encontrar fácilmente la densidad nutricional de lugares con los que son familiares. También sería de utilidad un tutorial interactivo guiado sobre como utilizar el mapa y comprender los indicadores que este expone. A nivel de administración del sitio, se podría facilitar un sitio de administración o algún método para subir y/o actualizar de forma sencilla los datos con los que este es alimentado, funcionalidad que no se implementó durante el desarrollo debido al enfoque en el desarrollo de las funcionalidades principales, así como el hecho de que los datos utilizados no cambiaron durante gran parte del desarrollo.

Por último, en cuanto a expansión de funcionalidades, se considera que la idea de integrar el modelo de densidad nutricional al sitio sería una excelente forma de exponer el trabajo realizado por el equipo de investigación de la USC. Esta funcionalidad requeriría habilitar un *endpoint* que permita traspasar los parámetros adecuados para que el modelo pueda realizar los cálculos predictivos necesarios y devolverlos al sitio. Dependiendo de la complejidad del modelo, es posible que esta funcionalidad requiera de un poder de procesamiento alto a nivel de backend.

Bibliografía

- [1] *Cartographic Boundary Files*. Disponible en <https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.2020.html>. Visitado por última vez: 29 de enero de 2024.
- [2] *Center for Disease Control And Prevention*. Disponible en <https://cdc.gov/>. Visitado por última vez: 29 de enero de 2024.
- [3] *Django Project*. Disponible en <https://www.djangoproject.com/>. Visitado por última vez: 29 de enero de 2024.
- [4] *Django REST Framework*. Disponible en <https://www.django-rest-framework.org/>. Visitado por última vez: 29 de enero de 2024.
- [5] *Food Access Research Atlas*. Disponible en <https://www.ers.usda.gov/data-products/food-access-research-atlas/>. Visitado por última vez: 29 de enero de 2024.
- [6] *Healthy Food Financing Initiative*. <https://www.investinginfood.com/>.
- [7] *PostGis*. Disponible en <http://postgis.net/>. Visitado por última vez: 29 de enero de 2024.
- [8] *React JS*. Disponible en <https://react.dev/>. Visitado por última vez: 29 de enero de 2024.
- [9] *Reinvestment Fund*. Disponible en <https://www.reinvestment.com/>. Visitado por última vez: 29 de enero de 2024.
- [10] *Southern California Center for Latino Health*. <https://health-equity-action.org/center/scclh>.
- [11] *User-Centered Design Basics*. Disponible en <https://www.usability.gov/what-and-why/user-centered-design.html>. Visitado por última vez: 29 de enero de 2024.
- [12] *What is REST?* Disponible en <https://restfulapi.net/>. Visitado por última vez: 29 de enero de 2024.
- [13] *¿Que es una API REST?* Disponible en <https://www.redhat.com/es/topics/api/what-is-a-rest-api>. Visitado por última vez 29 de enero de 2024.
- [14] Allcott, H., R. Diamond, J.P. Dubé, J. Handbury, I. Rahkovsky y M. Schnell: *Food deserts and the causes of nutritional inequality*. The Quarterly Journal of Economics, 134:1793 – 1844, 2019.
- [15] Devlin J., Chang M., Lee k. y Toutanova K.: *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018. <https://arxiv.org/abs/1810.04805>.
- [16] Douglas M Scheidt y Eileen Daniel: *Composite Index for Aggregating Nutrient Density Using Food Labels: Ratio of Recommended to Restricted Food Components*. 2004. <https://www.sciencedirect.com/science/article/abs/pii/S1499404606601267>.

- [17] *Food Environments LA*. Disponible en <https://foodenvironmentsla.dcc.uchile.cl/>. Visitado por última vez: 21 de diciembre de 2023.
- [18] Katz, S.F., F. Rodriguez y J.W. Knowles: *Health disparities in cardiometabolic risk among Black and Hispanic youth in the United States*. American Journal of Preventive Cardiology, 2021.
- [19] *Los Angeles Geohub*. Disponible en <https://geohub.lacity.org/>. Visitado por última vez: 29 de enero de 2024.
- [20] Liu, I., Abeliuk, A., de la Haye, K. y Horn, A.: *A continuous indicator of food environment nutritional quality*. 2021. <https://www.medrxiv.org/content/10.1101/2021.11.24.21266841v1>.
- [21] Norman, Donald A. y Stephen W. Draper: *User Centered System Design: New Perspectives on Human-Computer Interaction*. 1986.
- [22] Xu, Jiaquan, Sherry L. Murphy, Kenneth D. Kochanek y Elizabeth Arias: *Mortality in the United States, 2021*. NCHS Data Brief, 456, 2021. <https://www.cdc.gov/nchs/data/databriefs/db456.pdf>.

Anexos

Anexo A. Preguntas sobre usabilidad y eficiencia.

A continuación se presentan textualmente las preguntas realizadas a la investigadora de la USC, Abigail Horn, sobre la usabilidad, utilidad y eficiencia del prototipo desarrollado, así como sus respuestas.

1. Does the information provided by the map help in the identification of environments/restaurants with high or low nutritional density?

The map visualization is intended to be used by policymakers in public health and community-based organizations, as well as researchers conducting analyses into public health nutrition and environmental disparities. These users can use the visual to assess nutritional disparities and inform the allocation of resources for nutrition programming, such as identifying restaurants to include as Supplemental Nutrition Assistance Program (SNAP) (i.e., food stamps) authorized retailers, and in the Restaurant Meal Program, an LA County program that gives people that are at a high risk for chronic hunger the option to use CalFresh benefits to buy prepared meals at approved participating restaurant vendors.

The nutritional scores and visualization can also inform models for future implementation science efforts, and advance the ability of health advocacy groups through engagement to incentivize the food industry (restaurants and food delivery services) to provide healthier menus.

Yes, the information in the visualization helps these users to identify the locations of higher vs. lower nutrient density restaurants and food environments, helping to uncover areas of particularly low nutrient access that are not possible to identify with traditional category-based definitions of food environment nutritional quality. In addition, it helps these stakeholders to make planning assessments based on more accurate and refined information than the current standard, which is to assess nutritional quality of food environments based on the presence or absence of entire categories of food outlets (e.g., fast food, grocery).

2. Do you think that the statistical visualizations help to corroborate the existence or non-existence of relationships between the FEND indicator and the variables of interest?

It is currently hard to say whether the bivariate statistical results help to corroborate the existence/non-existence of relationships between the FEND indicator and the variables of interest, since the FEND data is not up to date. However, the statistical results are

very clearly visualized so as soon as the data are updated it should be possible to assess correlations between variables of interest. More comprehensive multivariable statistical models are ultimately needed to determine associations between the FEND indicator and variables of interest, but the bivariate correlations can provide an initial intuition for the direction of results.

3. Are the data delivered in a timely manner?

The data are delivered in a timely manner

4. Do you think that the data available for download has the necessary information for a third party to work with the indicators?

Almost, but it would be helpful to:

- Change the feature label “Namelsad” to “name” .
- Put all spatial variables next to each other (e.g., for census tract FEND: geo_type, tract, name).
- Round digits so as not to give false impression of precision.
- Include a download called README that explains the feature variables

5. What do you think would be good to add/remove/improve in future iterations of the project?

- More information to walk users through “how to use” the visualization, which could be done through interactive clicks or a visual tutorial
- Expand the portal to include a tool for users to input their own menu(s) to obtain nutrient composition and nutritional scores, allowing the application of the technology for estimating the nutritional quality of (a set of) restaurants that are not captured in the target food outlet menu database.
- Search the data for specific geographies, restaurants with particular names, or even menu item descriptors
- Highlight case studies around high vs low nutrient density neighborhoods
- Integrate with other used data resources like the USDA Food Environment Atlas scores and other neighborhood data