

# SABOR EN NÚMEROS

ANÁLISIS DE LOS  
RESTAURANTES EN ESPAÑA

Realizado por

**Kento Kamakura**  
**Pablo Lucas**  
**Miguel Gil**  
**Daniel Hernando**



**18-6-2024**

Proyecto: Comprensión de datos  
Ciencia de Datos-UPV



## ÍNDICE

<b>INTRODUCCIÓN.....</b>	<b>2</b>
<b>OBJETIVOS .....</b>	<b>2</b>
<b>NUESTROS DATOS: ADQUISICIÓN Y LIMPIEZA .....</b>	<b>3</b>
<b>PROCESO ADQUISICIÓN DE DATOS .....</b>	<b>4</b>
<b>PROCESO DE LIMPIEZA Y TRATAMIENTO DEL DATO .....</b>	<b>4</b>
<b>ALCANCE .....</b>	<b>7</b>
<b>ANÁLISIS EXPLORATORIO DE DATOS .....</b>	<b>8</b>
<b>ANÁLISIS UNIVARIANTE.....</b>	<b>9</b>
MICHELÍN .....	9
RESERVABLE .....	9
PRECIO MEDIO .....	10
RESERVAS ÚLTIMA SEMANA .....	10
RESEÑA DE COMIDA.....	11
<b>ANÁLISIS MULTIVARIANTE.....</b>	<b>12</b>
SALARIO MEDIO POR COMUNIDAD .....	12
PRECIO MEDIO DEL RESTAURANTE POR COMUNIDAD .....	12
TIPO DE COMIDA SEGÚN LA COMUNIDAD .....	13
CANTIDAD DE RESTAURANTES POR POBLACIÓN POR PROVINCIA .....	13
PRECIO MEDIO POR TIPO DE COMIDA .....	14
CORRELACIONES.....	15
PRECIO MEDIO Y SALARIO MEDIO .....	16
DISTINCIÓN Y VALORACIÓN DE COMIDA.....	16
PRECIO MEDIO Y VALORACIÓN DE COMIDA.....	16
NÚMERO DE PREMIOS Y PRECIO MEDIO.....	16
RESERVAS EN LA ÚLTIMA SEMANA CON RESTAURANTES MICHELÍN .....	16
PRECIO MEDIO Y RESTAURANTES MICHELÍN .....	16
SERVICIO Y AMBIENTE.....	17
RESERVAS EN UNA SEMANA Y VALORACIÓN DE COMIDA.....	17
NÚMERO DE FOTOS, DE COMENTARIOS Y DE VALORACIONES.....	17
RESTAURANTES POR POBLACIÓN Y SALARIO MEDIO ANUAL .....	17
<b>CONCLUSIONES AED .....</b>	<b>18</b>
<b>CRONOGRAMA.....</b>	<b>20</b>
<b>PROBLEMAS Y SOLUCIONES .....</b>	<b>21</b>
<b>CONCLUSIONES FINALES.....</b>	<b>22</b>
<b>MODELO PREDICTIVO.....</b>	<b>22</b>



## INTRODUCCIÓN

El presente proyecto titulado “Sabor en Números”, ha sido desarrollado para la asignatura Proyecto: Comprensión de Datos. Este trabajo colaborativo ha sido realizado por un equipo de cuatro estudiantes, cada uno con roles específicos para garantizar la eficiencia y calidad del proyecto.

El equipo está liderado por **Kento Kamakura**, quien ha desempeñado el rol de jefe del grupo, siendo el principal responsable de la gestión del tiempo y la coordinación de las entregas. **Miguel Gil** ha sido el encargado de la adquisición de datos, el desarrollo web y las predicciones. **Pablo Lucas** ha asumido la responsabilidad de garantizar la calidad y el tratamiento de los datos. Por último, **Daniel Hernando** ha sido el responsable del análisis exploratorio de los datos

En definitiva, el proyecto consiste en analizar las diversas características que influyen en el precio medio de un restaurante y desarrollar una solución tecnológica que pueda ayudar a los empresarios que desean emprender en el sector de la hostelería. A través de un análisis minucioso de los datos, el proyecto busca identificar patrones y tendencias que puedan ser útiles para futuros emprendedores.

## OBJETIVOS

Los objetivos específicos del proyecto son los siguientes:

- Conseguir una Base de Datos con los datos del precio medio por restaurante con un mínimo de 5000 datos.
- Conseguir una Base de Datos demográfica por provincias en España para cruzar los datos.
- Analizar qué características influyen más en el precio medio de un restaurante.
- Desarrollar un buscador de restaurantes dadas ciertas características del restaurante.
- Desarrollar un modelo de IA para la predicción del precio medio de un restaurante.
- Integrar todo el AED, Conclusiones y Soluciones tecnológicas en una única página web.



# Nuestros Datos: Adquisición y Limpieza

Sabor en números:  
Análisis de restaurantes en España

Proyecto: Comprensión de datos  
Ciencia de Datos - UPV

KENTO KAMAKURA GIMENO, PABLO LUCAS MORA, MIGUEL  
GIL JIMENEZ, DANIEL HERNANDO GRABOLEDA



## PROCESO ADQUISICIÓN DE DATOS

Una vez decidimos en grupo que queríamos analizar los datos de los restaurantes en España para encontrar relaciones con el precio medio, buscamos en Kaggle y otras plataformas de datos con el fin de descargar los datos directamente. Sin embargo, no encontramos ninguna base de datos que pudiese satisfacer nuestras necesidades. Por ello, se nos ocurrió aplicar técnicas de webscraping a The Fork, con el objetivo de hacernos con los datos de todos sus restaurantes reales. Aunque esta tarea nos llevase algo más de tiempo, éramos conscientes de la necesidad de realizarlo por la falta de datos y para asegurarnos la veracidad de los datos. Para ello creamos un software con Python haciendo uso de librerías como Selenium, Requests, BeautifulSoup y Pandas. Este software se puede encontrar en nuestro repositorio de GitHub público.

Además, aprovechamos el potencial de Python en este proceso de adquisición de datos para ir limpiando algunos datos durante la propia adquisición de datos como puede ser eliminar el símbolo “€” de la variable precio medio o convertir la variable “Estrella Michelin” en un booleano, cuando originariamente era un campo nulo o una url de la estrella michelín.

Una vez completado este proceso de adquisición de datos de The Fork procedimos a convertirlo en un CSV para facilitar su envío y posterior tratamiento de los datos.

Por otro lado, consideramos interesante inicialmente conseguir una base de datos de Población, Salario medio y Asalariados por provincia tal y como establecimos en nuestros objetivos. Habiendo adquirido estas 3 bases de datos del INE, procedimos a unir las 4 base de datos finalmente en un CSV con el fin de comenzar a limpiar y posteriormente explorar los datos.

## PROCESO DE LIMPIEZA Y TRATAMIENTO DEL DATO

Una vez ya teníamos la base de datos, pasamos a la limpieza y el tratamiento de los datos. Para ello fuimos columna por columna haciendo una pequeña previsualización de los datos que se obtenían en cada variable (análisis univariante). Al realizar este análisis nos dimos cuenta de que algunas de ellas tenían datos faltantes o anómalos, que debíamos revisar para ver qué tratamiento les dábamos a cada caso.

Uno de los primeros datos faltantes que observamos fue en la columna de ‘Rate\_distinction’, en la cual aparecían los restaurantes que tenían algún tipo de distinción (Excellent, Fabulous o Very Good). Esta variable tenía una gran cantidad de datos faltantes, pero tras estudiarlos observamos que no eran errores, ya que no todos los restaurantes tenían porque tener distinción. Por tanto, como estos faltantes aportaban un valor a la variable, los dejamos en la columna. Además, para un más fácil análisis de estos datos, decidimos hacer una asignación





de un número a cada distinción, por su orden de magnitud (1 Very Good, 2 Fabulous y 3 Excellent).

El siguiente cambio que realizamos fue la transformación de la variable 'Michelin' (cuyos valores eran booleanos) a valores numéricos (1 True y 0 False), pero solo con el objetivo de facilitar el análisis, no por algún tipo de fallo.

Al proseguir con nuestro análisis univariante, en la columna de 'Provincia', encontramos que había más valores únicos que provincias en España, lo que nos hizo analizar con más detenimiento lo que sucedía. Observamos que, en algunos, The Fork había introducido el nombre del pueblo o ciudad a la que pertenecían en vez de la provincia. Como en este caso se trataba de pocos datos erróneos, los decidimos corregir manualmente. Para ello, buscamos en Google Maps el nombre de la provincia a la que pertenecía el pueblo o ciudad, y lo imputamos.

Posteriormente, observamos que en las columnas de precios medios y de tipo de comida por cada restaurante había algunos datos faltantes, pero al tratarse de muy pocos (menos de 90) decidimos introducir los datos manualmente mediante la búsqueda en internet de estos mismos (Precio medio aportado por Google).

Continuando con el análisis de todas las columnas, después de la unión de las bases de datos de salarios medios, asalariados y población por provincia con nuestra base inicial vimos que no estaban los datos de Navarra y País Vasco de asalariados y sueldo medio. Al encontrarnos con este problema decidimos buscar estos por internet, e implementarlos manualmente. En el caso de los salarios medios anuales por provincia encontramos fuentes fiables y contrastadas en distintas páginas, por lo que los introducimos en la base de datos. En cambio, al buscar los datos de asalariados manualmente por internet, no fuimos capaces de encontrar fuentes fiables a diferencia de los anteriores. Y los de población no nos dieron ningún tipo de problema, ya que estaban completos y eran fiables. Por lo que finalmente decidimos quedarnos solo con la columna de salarios medios y población.

Para cerrar, tras una última revisión decidimos eliminar los datos en los cuales el precio medio era menor de 6. Esta decisión la tomamos debido a que significaban una fracción muy pequeña de la población (4 datos), y que, además, analizar uno por uno nos dimos cuenta de que se debía a un error de The Fork, como por ejemplo el caso de un restaurante en el cual aparecía un precio medio de 1 euro y que tras mirarlo manualmente en The Fork observamos



que se debía a que solo aparecían los precios de las bebidas y no los de los platos (Por lo que The Fork realiza la media solo de los precios de las bebidas).

<b>Entrante</b>	
Croquetas caseras	
Anchoas Santoña	
Ventresca de Bonito con pimientos	
<b>Plato</b>	
Bacalao en salsa de manitas	
<b>Postre</b>	
Arroz con leche	
<b>Bebidas</b>	
Botella de agua .....	2 €
Media botella de agua .....	1,5 €
Café .....	1,5 €

Además, observamos que había algunos restaurantes con 0 métodos de pago, cosa que es imposible ya que tiene que existir alguna manera para pagar, por lo que hemos decidido prescindir de estos, ya que está claro que son datos erróneos. Como eran pocos datos se decidió eliminar estas filas.



## ALCANCE

1. MISIÓN	-Analizar las características que influyen en el precio medio de un restaurante -Aportar una solución tecnológica a empresarios en el sector hostelería.
2. REQUISITOS	-AED detallado (Statgraphics y Python), Control de versiones git, No Powerpoint
3. RESTRICCIONES	-Entrega en junio -Encontrar base de datos con mínimo 5000 restaurantes
4. SUPUESTO INICIAL	-Información en <b>The Fork</b> : Actualizada y verídica (sin conflicto de intereses) -La muestra de los restaurantes de The Fork es representativa a nivel España
6. ENTREGABLES	-Informe del alcance, cronograma, Informe adquisición y limpieza, Informe AED -Web y Repositorio: <a href="#">webscraping</a> , limpieza de datos, código de la web
7. LÍMITES	-No entramos en analizar platos ni menús. -La web sugerirá un precio medio para un restaurante, pero no el tipo de comida o localización del mismo.
9. CRITERIOS ÉXITO	-El AED confirma si existe o no una relación con el precio medio del restaurante -La solución tecnológica es útil para los empresarios en el sector hostelería





# Análisis Exploratorio de Datos

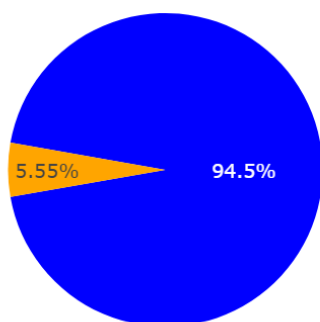
Sabor en números:  
Análisis de restaurantes en España

Proyecto: Comprensión de datos  
Ciencia de Datos - UPV

KENTO KAMAKURA GIMENO, PABLO LUCAS MORA, MIGUEL  
GIL JIMENEZ, DANIEL HERNANDO GRABOLEDA

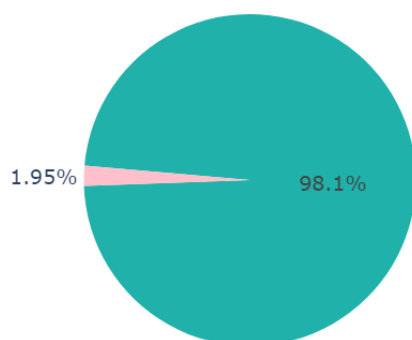
# ANÁLISIS UNIVARIANTE

## MICHELÍN



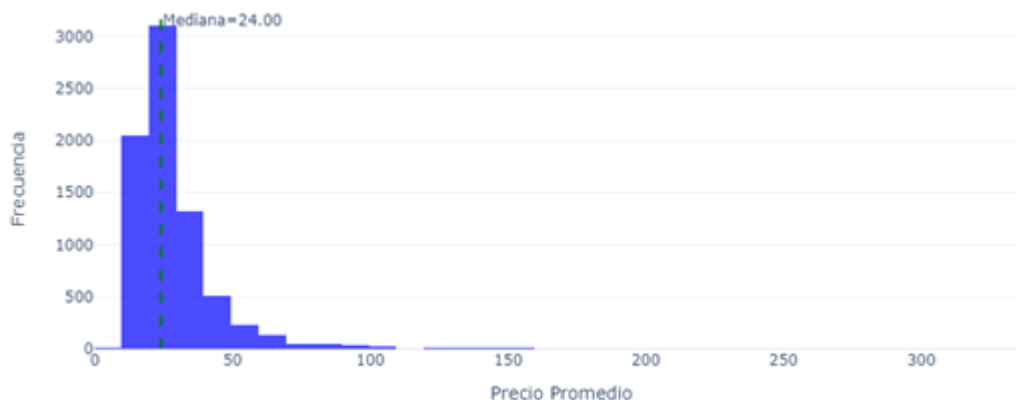
Como vemos, solo el 5'55% de los restaurantes tiene una o más estrellas michelín. Esto demuestra en gran parte la dificultad que conlleva conseguir una estrella michelín en España

## RESERVABLE



Aquí se muestra que actualmente en España el servicio de reserva es prácticamente del 98%, resaltando que a la gran mayoría de los restaurantes se puede acudir mediante la reserva previa de la mesa.

## PRECIO MEDIO



En primer lugar, eliminamos las columnas cuyo precio medio era menor a 6 euros y es que por muy barato que sea el restaurante, no es lógico que en España puedas comer por 6 euros. El precio medio por comensal según el restaurante es otra variable muy dispersa y tiene un rango de 307 valores, indicando la presencia de muchos valores distintos y un comportamiento variado como vemos en el diagrama de barras. Por ello decimos que presenta una asimetría positiva debido a una mayor concentración de valores al principio y la presencia de valores atípicos como 307, que hacen que la medida adecuada de centralización sea la mediana, cuyo valor, 24, es mucho más razonable. Por último, es muy reseñable la curtosis de 632, que es mucho mayor a 0, evidenciando una forma muy leptocúrtica.

## RESERVAS ÚLTIMA SEMANA





Esta es de las variables más dispersas y tiene un coeficiente de variación del 229%, junto con un rango de 1133 valores, indicando la presencia de muchos valores distintos y un comportamiento irregular como el diagrama de barras. Es por ello por lo que presenta una asimetría positiva debido a una mayor concentración de valores al principio y la presencia de valores atípicos como 1133, que hacen que la medida adecuada de centralización sea la mediana o media recortada. Por último, es muy reseñable la curtosis de 1014, que es mucho mayor a 0, evidenciando una forma muy leptocúrtica.

## RESEÑA DE COMIDA



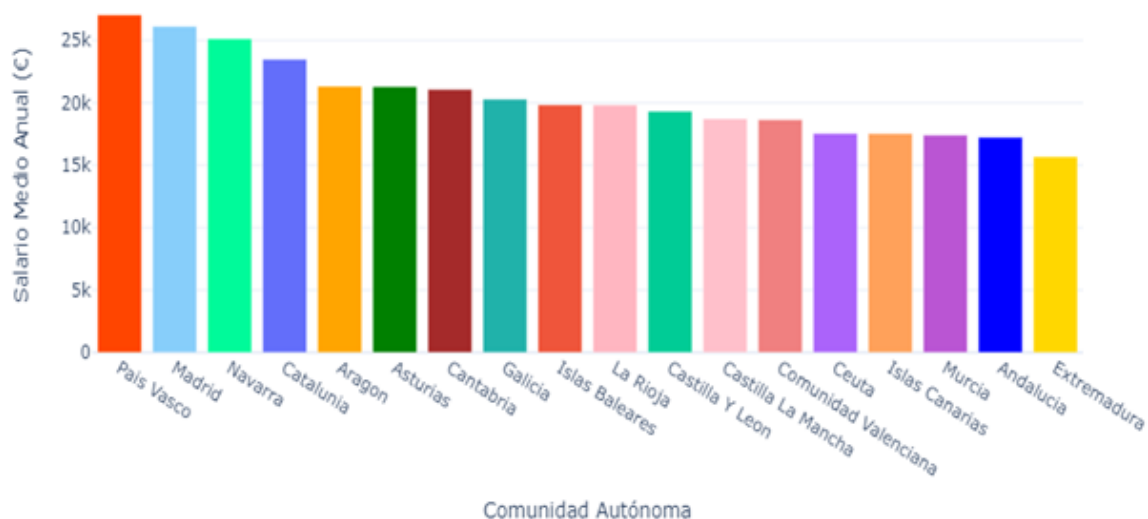
En la valoración de la comida, observamos que el rango de valores no es extenso y que todos se concentran entre el 4 y el 10, siendo los números más elevados los con mayor presencia de valores, así que presenta una asimetría negativa por una mayor concentración de valores al final y una menor al principio.



## ANÁLISIS MULTIVARIANTE

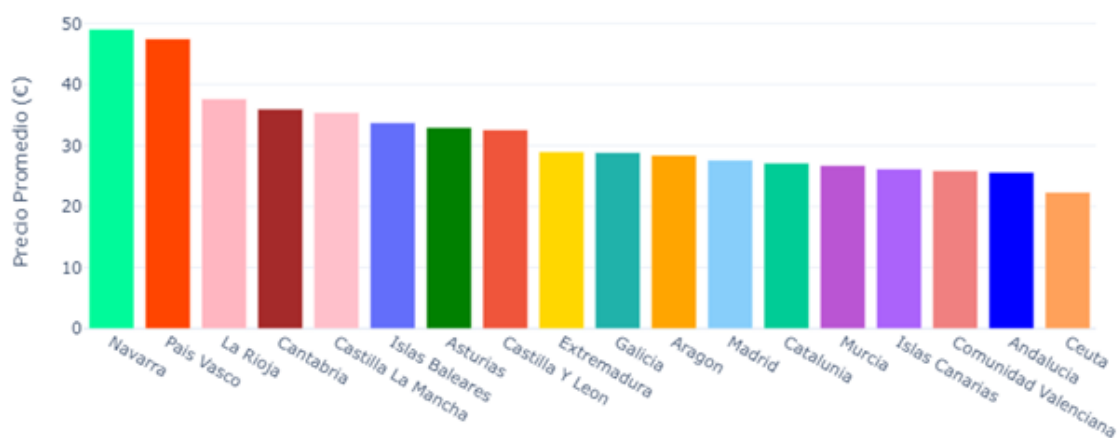
A continuación, mostramos las siguientes gráficas que recogen los salarios y precios medios según la ubicación, que nos guiarán a la hora de sacar conclusiones

### SALARIO MEDIO POR COMUNIDAD



En primer lugar, observamos el sueldo medio de los españoles según la comunidad autónoma, liderado por País Vasco y Madrid y con Extremadura como comunidad con el salario más bajo.

### PRECIO MEDIO DEL RESTAURANTE POR COMUNIDAD



La siguiente gráfica muestra por orden el precio medio por persona de una comida por



comunidad. Esto nos revela que la comunidad más cara es Navarra, con un precio medio de casi 50€, frente a Ceuta, donde apenas sobrepasa los 20€.

### TIPO DE COMIDA SEGÚN LA COMUNIDAD

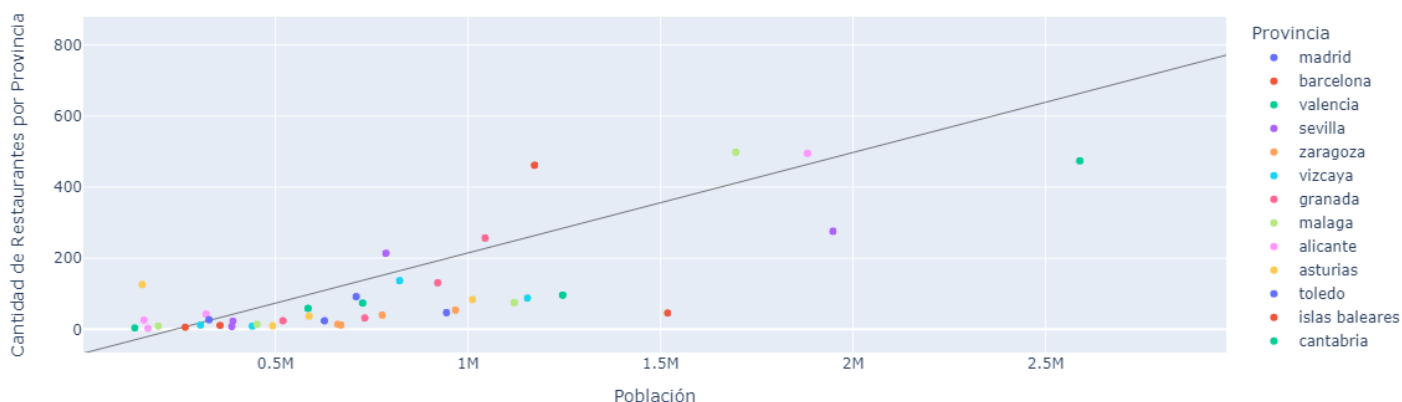
COMUNIDAD AUTÓNOMA	COMIDA MÁS CONSUMIDA 1	COMIDA MÁS CONSUMIDA 2	COMIDA MENOS CONSUMIDA 1	COMIDA MENOS CONSUMIDA 2
Andalucía				
Galicia				
C. Valenciana				
Madrid				
Cataluña				
Islas Canarias				
Asturias				

Hemos recogido una tabla que muestra los tipos de comida más repetidos en las siguientes comunidades:

Se puede concluir como en las grandes ciudades aparece más tipo de comida internacional, debido al tráfico de turistas y extranjeros, siendo la comida española y sobre todo la mediterránea la principal. Por otro lado, en las comunidades más pequeñas y de menor población predomina la comida local. A pesar de ello, las comunidades con fuerte identidad gastronómica suelen superar al resto (véase C. Valenciana, Cataluña, Andalucía...)

No obstante, el tipo de comida menos consumido en las diferentes comunidades autónomas son aquellos alimentos propios de la cocina internacional menos representada, como la comida coreana, india, venezolana, peruana...

### CANTIDAD DE RESTAURANTES POR POBLACIÓN POR PROVINCIA

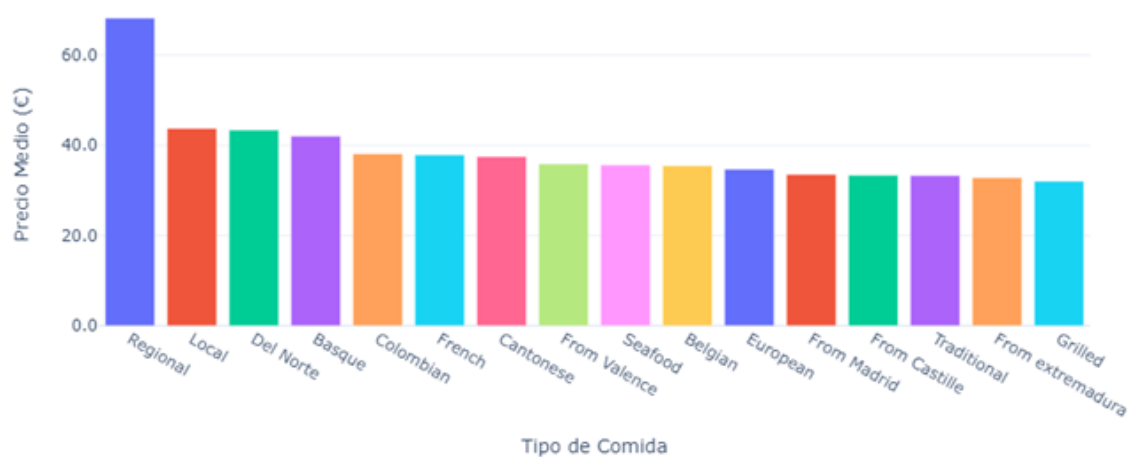


Este gráfico de puntos recoge los datos de la cantidad de restaurantes por provincias. La línea de tendencia muestra que estas variables siguen una relación proporcional y totalmente lineal, es decir, cuanta más población hay en una provincia, más cantidad de restaurantes posee

## PRECIO MEDIO POR TIPO DE COMIDA

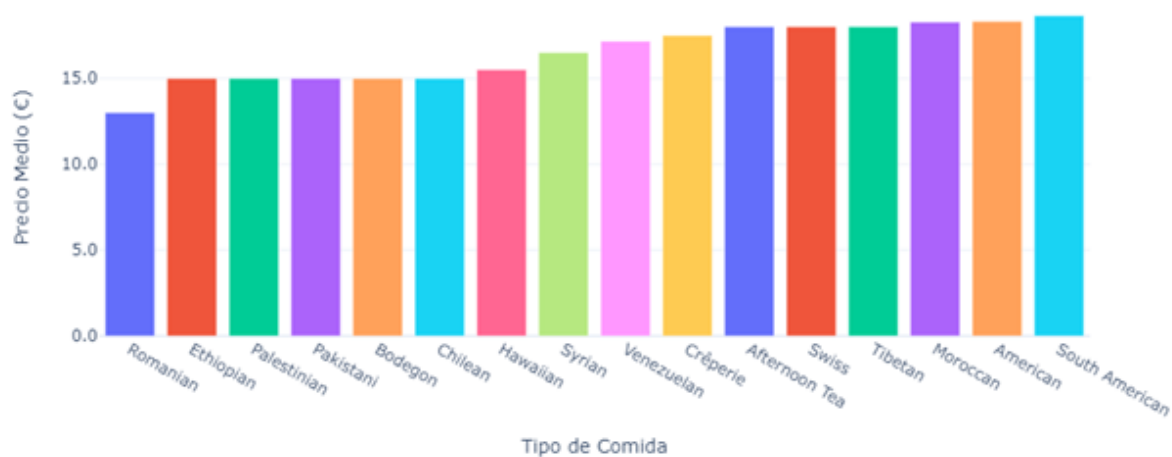
En los siguientes gráficos de barras hemos recogido los principales tipos de comida y los hemos representado según su precio medio:

- **Mayor a menor precio**



Como vemos, la comida local y del norte de España, sobre todo del País Vasco, es la más cara.

- **Menor precio**







Mientras que la comida proveniente de países como Etiopía, Palestina, Pakistán y Rumanía es la más barata.

## CORRELACIONES

Aquí mostramos las diferentes correlaciones de Pearson estudiadas en nuestro proyecto mediante la siguiente tabla, evitando así poner el gráfico de correlaciones de Pearson en cada una de las relaciones, y es que sólo nos interesa el número, por lo que simplificamos el resultado de esta forma:

<b>Variables estudiadas</b>	<b>Correlación de Pearson</b>
Precio medio – salario medio	0,03
Distinción - Valoración de comida	0,83
Precio medio - Valoración de comida*	0,22
Número de premios - Precio medio	0,55
Reservas en una semana - Restaurantes Michelin*	0,07
Estrella Michelin - Precio medio	0,59
Servicio - Ambiente	0,79
Reservas en una semana - Valoración de comida	0,09
Número de comentarios - Número de fotos	0,1
Número de fotos - Valoraciones	0,1
Número de comentarios - Valoraciones	0,98
Población - Cantidad de restaurantes	0,91
Salario medio - Cantidad de restaurantes	0,32
Salario medio - Población	0,43



A continuació, explicamos de forma més detallada el significado del número obtenido en cada relación:

## PRECIO MEDIO Y SALARIO MEDIO

Se observa que la relación precio medio y salario medio es prácticamente nula y puede estar influenciada por una variedad de factores adicionales, como la diversidad económica, la demanda y la oferta, los costos operativos y la segmentación del mercado.

## DISTINCIÓN Y VALORACIÓN DE COMIDA

La distinción de un restaurante y la valoración de su comida están estrechamente relacionadas, ya que los estándares de calidad, innovación, ingredientes y consistencia en la oferta gastronómica son elementos clave que influyen tanto en la obtención de reconocimientos como en la satisfacción de los comensales.

-

## PRECIO MEDIO Y VALORACIÓN DE COMIDA

Llama también la atención la escasa relación entre el precio medio y la valoración, pues se tiende a asociar una buena valoración con mayor precio del restaurante. Este gráfico de correlaciones Pearson demuestra que no hay mucha conexión, por lo que mayor precio medio no implica una mejor comida y viceversa.

## NÚMERO DE PREMIOS Y PRECIO MEDIO

En cuanto a las variables Numero\_awards (número de premios del sitio) y Precio Medio, observamos una ligera conexión, por lo que concluimos que hay cierta relación entre los precios sobre los que suele rondar un restaurante para que obtenga premios.

## RESERVAS EN LA ÚLTIMA SEMANA CON RESTAURANTES MICHELÍN

Esta se trata de la correlación que más se salió de nuestras expectativas al realizar el análisis multivariado de correlaciones, dado que los restaurantes con estrella Michelin son muy populares por su calidad y reciben un muy elevado número de reservas por semana. Una vez nos pusimos a investigar a qué se debe esto, llegamos a la conclusión de que las reservas en este tipo de restaurantes se realizan a través de sus propias páginas web, y no mediante The Fork. Esto explicaría por qué en nuestra base de datos una gran cantidad de estos restaurantes no reciben reservas.

## PRECIO MEDIO Y RESTAURANTES MICHELÍN

Como era de esperar, comer en un restaurante con estrellas Michelin suele ser considerablemente más caro que en otros tipos de restaurantes. Los comensales pagan no solo por la comida, sino por la experiencia culinaria completa, que incluye servicio, ambiente y la innovación en cada plato.



## SERVICIO Y AMBIENTE

El servicio y el ambiente de un restaurante están fuertemente relacionados y se complementan entre sí para crear una experiencia gastronómica más satisfactoria para los comensales. Un ambiente acogedor y bien diseñado puede realzar la calidad del servicio, mientras que un servicio amable y atento puede mejorar la percepción del ambiente por parte de los clientes.

## RESERVAS EN UNA SEMANA Y VALORACIÓN DE COMIDA

Para nuestra sorpresa, el número de reservas en la última semana no tiene que ver con la valoración de comida. Esto quiere decir que, porque un restaurante esté muy reservado o frecuentado, no está comprometida su calidad.

## NÚMERO DE FOTOS, DE COMENTARIOS Y DE VALORACIONES

En este gráfico analizamos tres variables: número de valoraciones, de comentarios y la cantidad de fotos subidas del restaurante.

A pesar de no hallar relación entre el número de fotos y las valoraciones y/o comentarios, el número de comentarios y valoraciones sí que presenta una relación lineal, pues los clientes que dejan un comentario dejan seguidamente una puntuación sobre el sitio donde han consumido.

## RESTAURANTES POR POBLACIÓN Y SALARIO MEDIO ANUAL

En el siguiente gráfico observamos las correlaciones que existen entre la cantidad de restaurantes por cada comunidad, la población de esta, y su salario medio anual. Al observar los valores que nos da, destaca la fuerte relación que existe entre la población y el número de restaurantes, cosa que si nos paramos a pensar tiene mucho sentido ya que cuanto mayor número de habitantes, más clientes y necesidades de comida, y por tanto más restaurantes. Esta relación, además de por la correlación de Pearson, también la evidenciamos anteriormente mediante un gráfico de puntos en el que vimos la proporcionalidad entre los restaurantes y la población. Y por otro lado vemos algo de relación entre la población y el salario medio, y el salario medio y la cantidad de restaurantes, pero al no superar el 0,5 no consideramos que sea un valor representativo para definir sus relaciones.



## CONCLUSIONES AED

El objetivo básico planteado al inicio del trabajo fue comprobar qué es lo que influye en el precio de un restaurante en España y saber si influyen ciertos factores como la localización, la renta per cápita, el tipo de comida, etc. Para ello, se han analizado los resultados de 7624 restaurantes de las diferentes provincias españolas y se han establecido diferencias y relaciones entre ellas que nos han ayudado a tomar ciertas conclusiones y de algún modo, 'solucionar' el problema.

En primer lugar, consultamos los datos de las comunidades autónomas con mayor salario medio anual, y mediante el uso de gráficas hacernos una mejor idea de dónde se sitúa cada comunidad de España, observando que el sueldo medio de los españoles está liderado por País Vasco y Madrid y con Extremadura como comunidad con el salario más bajo.

De la misma forma, consultamos en nuestra base de datos los tipos de comida según su precio y los ordenamos para facilitar nuestras conclusiones en el futuro.

Con la variable **precio medio**, un parámetro fundamental a la hora de elegir un restaurante, encontramos diversas relaciones. En primer lugar, con el salario medio, encontrando una escasa relación, influenciada por una variedad de factores adicionales. Continuando con la variable comunidad: la comunidad más cara es Navarra y el País Vasco, con un precio medio de casi 50€, frente a Ceuta, donde apenas sobrepasa los 20€. Esto tiene mucho que ver con el precio medio según el tipo de comida, y es que la comida local y del norte de España (sobre todo del País Vasco) es la más cara. También es destacable la escasa relación que tiene con la valoración, pues se tiende a asociar una buena valoración con un mayor precio del restaurante, esto se podría explicar debido a que la gente no solo busca mucha calidad (altos precios), sino que busca una buena relación calidad precio, es decir, poder permitirse una buena comida a un precio razonable y asequible. Sí tiene más relación con el número de premios del restaurante, y es que un restaurante con muchos premios suele ser más caro. Lo mismo podemos concluir con la variable estrella Michelin, indicando que, como norma general, si un restaurante tiene estrella Michelin significa que el precio de éste será mayor.

En cuanto a la **valoración de la comida**, este se trata de otro aspecto básico a la hora de elegir un restaurante, pues ofrece una visión importante sobre uno de los aspectos críticos de la experiencia en un restaurante y puede influir significativamente en la percepción general del establecimiento por parte de los clientes potenciales. Por un lado, al relacionarlo con la distinción, encontramos una correlación entre ambas. No obstante, para nuestra sorpresa, el número de reservas en la última semana no tiene que ver con la valoración de comida. Esto quiere decir que, porque un restaurante esté muy reservado o frecuentado, no se encuentra comprometida su calidad.

Respecto al **número de comentarios** en un restaurante, esta es una variable que nos puede indicar su popularidad, relevancia, calidad percibida, credibilidad, etc, y sí que presenta una relación lineal con el número de valoraciones, pues los clientes que dejan un comentario dejan seguidamente una puntuación sobre el sitio donde han consumido. No obstante, el número de comentarios no implica un mayor número de fotos en la web, y es que la relación es prácticamente nula, como ya demostramos en su correspondiente apartado.

Por otro lado, también se evidencia la conexión entre las valoraciones del servicio y el ambiente, un ambiente acogedor y bien diseñado puede realzar la calidad del servicio,



mientras que un servicio amable y atento puede mejorar la percepción del ambiente por parte de los clientes, lo que demuestra la importancia del ambiente que te rodea y el sentirse a gusto en un restaurante.

Según el **tipo de comida**, en relación con la comunidad, se puede concluir como en las grandes ciudades aparece más tipo de comida internacional, debido al tráfico de turistas y extranjeros, siendo la comida española y sobre todo la mediterránea la principal. Por otro lado, en las comunidades más pequeñas y de menor población predomina la comida local. A pesar de ello, las comunidades con fuerte identidad gastronómica suelen superar al resto. Por otro lado, el tipo de comida menos consumido en las diferentes comunidades autónomas son aquellos alimentos propios de la cocina internacional menos representada, como la comida coreana, india, venezolana, peruana...

Por último, contemplamos las correlaciones que existen entre la cantidad de restaurantes por cada comunidad, la población de esta, y su salario medio anual. Al observar los valores que nos da, destaca la fuerte relación que existe entre la población y el número de restaurantes, algo que tiene mucho sentido, ya que cuanto mayor sea el número de habitantes, más clientes y necesidades de comida, y por tanto más restaurantes. Finalmente, vemos algo de relación entre la población, el salario medio y la cantidad de restaurantes, pero al no superar el 0,5 no consideramos que sea un valor representativo para definir sus relaciones.

## CRONOGRAMA

En el ecuador de nuestro proyecto, el cronograma reflejaba claramente nuestro progreso y el cumplimiento de los tiempos establecidos. Este cronograma nos permitió organizar de manera eficiente nuestras tareas y mantener una visión del avance global del proyecto

OBJETIVOS PRINCIPALES	OBJETIVOS ESPECÍFICOS	26-feb	4-mar	11-mar	18-mar	25-mar	1-abr	8-abr	15-abr	22-abr	29-abr	6-may	13-may	20-may	27-may
ADQUISICIÓN DE DATOS	Queda BBDD Población por Provincia	Pablo, Daniel													
	Queda BBDD de Salarios por Provincia														
	Webscrapping a The Fork	Miguel													
TRATAMIENTO DE DATOS	Limpieza de Datos		Todos												
	Unión Bases de datos														
AED	AED Univariente			Pablo, Kento, Daniel					Pablo, Kento, Daniel						
	AED Multivariente				Pablo, Kento, Daniel										
Desarrollo Web	Maquetación web inicial			Miguel											
	Decisión de gráficas en web									Pablo, Kento, Daniel					
	Pasar las gráficas a Python									Pablo, Kento, Daniel					
	Implementar las gráficas en la web							Kámen			Miguel				
	Crear buscador y dashboard de restauración				Miguel				Miguel						
	Hacer el análisis predictivo														
Entregas, control Revisión	Forma de adquisición y tratamiento de datos														
	Crear informe del AED								Pablo						
	Control de versiones de GitHub	Miguel	Miguel	Miguel	Miguel	Miguel	Miguel		Miguel	Miguel	Miguel	Miguel	Miguel	Miguel	Miguel
	Crear Cronograma		Todos												
	Crear PDF de la entrega final												Todos		
	Crear video de la entrega														
	Entrega Semanal	Daniel	Daniel	Daniel	Daniel	Daniel	Daniel		Daniel	Daniel	Daniel	Daniel	Daniel	Daniel	Daniel
	Crear Presentación de los avances								Todos						
ESTADOS								Hasta 16:							
RETRASADO								Entrega una presentación							
EN PROCESO								a realizar el							
POR HACER								17 o 18.							
FINALIZADO								Duración: 15							

Actualmente, hemos logrado completar todas las tareas impuestas inicialmente y nos encontramos en la fase final, enfocándonos en retoques y optimizaciones necesarias para perfeccionar el trabajo.

OBJETIVOS PRINCIPALES		OBJETIVOS ESPECÍFICOS	26-feb	4-mar	11-mar	18-mar	25-mar	1-abr	8-abr	15-abr	22-abr	####	6-may	13-may	20-may	27-may
ADQUISICIÓN DE DATOS	Queda BBDD Población por Provincia	Pablo, Kento, Daniel														
	Queda BBDD de Salarios por Provincia	Miguel														
TRATAMIENTO DE DATOS	Web scraping a The Fork															
	Limpieza de Datos			Todos												
	Unión Bases de datos															
AED	AED Univariante				Pablo, Kento, Daniel											
	AED Multivariante					Pablo, Kento, Daniel				Pablo, Kento, Daniel						
Desarrollo Web	Maquetación web inicial				Miguel											
	Decisión de gráficas en web										Pablo, Kento, Daniel					
	Pasar las gráficas a Python										Pablo, Kento, Daniel					
	Implementar las gráficas en la web											Miguel				
	Buscar y dashboard de restaurantes						Miguel									
	Hacer el análisis predictivo										Miguel					
Entregas, control Revisión	Forme de adquisición y tratamiento										Pablo					
	Crear informe del AED										Pablo, Kento, Daniel					
	Control de versiones de GitHub	Miguel	Miguel		Miguel	Miguel	Miguel	Miguel		Miguel	Miguel	Miguel	Miguel	Miguel	Miguel	Miguel
	Crear Cronograma		Todos													
	Crear PDF de la entrega final														Todos	
	Crear video de la entrega															
	Entrega Semanal	Daniel	Daniel		Daniel	Daniel	Daniel	Daniel		Daniel	Daniel	Daniel	Daniel	Daniel	Daniel	Daniel
	Crear Presentación de los avances									Todos						
ESTADOS									Hasta 16:							
RETRASADO									Entregar una							
EN PROCESO									presentación							
POR HACER									n a realizar el							
FINALIZADO									17 o 18.							
									Duración:							



## PROBLEMAS Y SOLUCIONES

### 1. Datos faltantes Navarra y País Vasco

- Faltaban sus datos de salarios medios y asalariados.
  - Hicimos una búsqueda internet.
- Encontramos datos fiables para los salarios medios.
  - No conseguimos los datos de asalariados.
- Imputamos los salarios medios obtenidos manualmente.

### 2. Faltantes en la unión de bases de datos

- Había variables que contaban con muchos datos faltantes, pero como estos tenían un significado los hemos mantenido.
- Había otras con pocos que corregimos manualmente.
  - Como los de precio medio.

### 3. Provincias mal puestas

- En la columna de provincias había algunas que estaban mal puestas en [The Fork](#) (había algún nombre de pueblo o ciudad en vez del de la provincia).
- Como se trataba de pocos datos erróneos los corregimos manualmente.
  - Mediante una búsqueda en [Google Maps](#).





## CONCLUSIONES FINALES

En el marco de nuestro proyecto sobre la gastronomía en España, hemos alcanzado todos los objetivos planteados inicialmente. En primer lugar, conseguimos crear una base de datos de precio medio por restaurante con más de 5000 datos mediante web scraping y otra base de datos demográfica por provincias en España a través del INE.

Mediante un exhaustivo análisis exploratorio de los datos, logramos sacar relaciones entre nuestras variables y pudimos comprender el panorama de los restaurantes en España, analizando los precios medios, los tipos de comida más demandados y las distintas reseñas que dejan los usuarios, para así implementarlo en nuestra web. No solo hemos logrado esto, sino que también hemos implementado un buscador de restaurantes basado en características geográficas, económicas y tipos de comida. Además, hemos integrado un modelo predictivo que estima el precio medio de un restaurante.

En definitiva, hemos cumplido y superado satisfactoriamente los objetivos iniciales del proyecto, creando una página web completa y atractiva en la que identificamos patrones y tendencias que puedan ser útiles para los futuros empresarios que desean emprender en el sector de la hostelería y que también ayuden al cliente a elegir un restaurante según sus necesidades.

Para más información sobre nuestro proyecto y explorar la plataforma, te invitamos a visitar nuestra página web: <https://thefork.streamlit.app/>

## MODELO PREDICTIVO

El último avance que introdujimos en nuestra web fue un modelo predictivo que tiene el fin de predecir el precio medio de un restaurante conociendo sus características principales. Finalmente, desarrollamos 3 modelos diferentes. Un Linear Regression, un Random Forest y un XGBoost y hacemos uso de un Voting Regressor. Esto significa que la predicción final será la media de la predicción de cada uno de los modelos. Sin embargo, no conseguimos que los resultados proporcionados por el modelo se acerquen a la realidad y consideramos que habría que mejorar estos modelos bastante antes de poder ponerlo en producción realmente. Consideramos que no lo hemos conseguido perfeccionar por una combinación de falta de conocimiento técnico en cuanto a modelos predictivos y por la aleatoriedad que encontramos en los datos tal y como se menciona en el AED. Consideramos que nuestro modelo puede estar cometiendo underfitting y presenta las siguientes métricas de errores Mean Squared Error: 256.45 y Root Mean Squared Error: 16.01. Por último, consideramos que esta podría ser una propuesta a futuro para mejorar el proyecto, que va más allá de la asignatura actual.