

## *Exploratory Data Analysis*

*Travelling around the  
World*

----- Álvaro Martinez -----

by Next Gen Travel

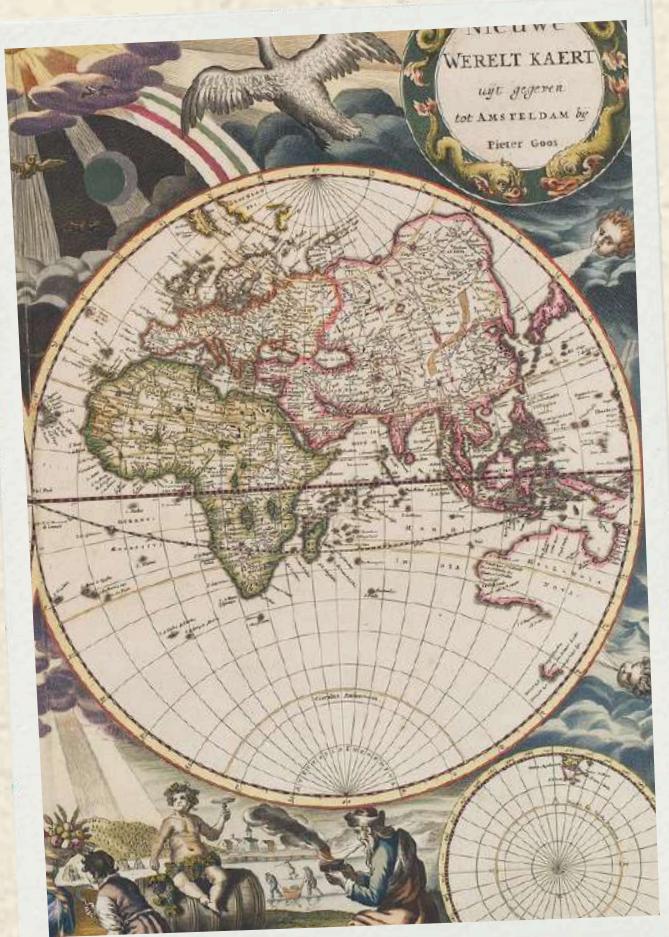




# íNDICE

---

- 01** ¿Quienes somos?
  - 02** Contexto, hipótesis y opciones de negocio.
  - 03** Making off:  
recopilación de datos  
y estructura
  - 04** Análisis Univariante
  - 05** Análisis por categoría
  - 06** Conclusions
- 



# ¿Quienes somos?

---

**NexGen Travel** se define como la primera consultora turística del mundo, ofreciendo soluciones adaptadas y personalizadas a nuestros clientes, ya sean turistas, profesionales del sector turístico e incluso países.

En **NextGen Travel** conocemos a nuestros clientes, y sabemos que un destino no son solo playas o monumentos, sino una combinación de muchos factores que tenemos en cuenta para hacer la experiencia inolvidable.

En **NextGen Travels** nos apasionan los datos, sabemos que hacer con ellos y traducir DataFrames en experiencias para los turistas, oportunidades de negocio para profesionales e incluso recomendaciones para mejorar el turismo en países.

De esta forma modernizamos el sector turístico ofreciendo a cada uno lo que necesita y consiguiendo que todos seamos un poquito más felices en este mundo.

# Contexto, hipótesis y negocio

## 01. Turistas vs Realidad

Este es un ejemplo de EDA que analiza opiniones de turistas en 17 países basados en múltiples categorías como gastronomía, seguridad, cultura, paisajes

Paralelamente se recuperaron datos reales de los países relacionados con las mismas categorías que los turistas califican.

Nuestro objetivo es saber en qué categorías los turistas tienen una visión realista de lo que pasa en los países, o si por el contrario es una opinión sesgada basada en una “burbuja” vacacional

Esto nos puede ayudar a conocer más a nuestros turistas y saber qué destinos ofrecerles, a la vez que a los profesionales del turismo les ayude a ofrecer mejores servicios.

## 02. ¿Donde invierto?

También somos una empresa de asesoramiento a profesionales y ofrecemos soluciones de negocio a preguntas reales.

En este caso damos una pincelada de lo que podemos ofrecer haciendo un análisis rápido de cuál sería el destino idea para poner un hotel basandonos en los datos recopilados.



# Making off: ¿Como hemos llegado aqui?



# Making off: Los Metadatos

24 webs consultadas

17 Países

2 extracciones Selenium

7 extracciones CSV

2 extracciones API

3 extracciones xlsx

7 extracciones JSON

17 extracciones Web Scraping

Latinoamérica: México,  
Colombia, Perú, Bolivia, Chile,  
Brasil, Argentina

Sudeste Asiático: Malasia,  
Laos, Filipinas, Camboya,  
Tailandia, Vietnam, Indonesia

Oceanía: Australia,  
Nueva Zelanda

Asia Oriental: China

Dataframe: 17 filas x 52 columnas  
(884 datos)

Variables

Opinión de turistas (13 categorías)

Economía (6 categorías)

Gastronomía (4 categorías)

Seguridad y Criminalidad (4 categorías)

Amabilidad y Facilidad de Adaptación  
(5 categorías)

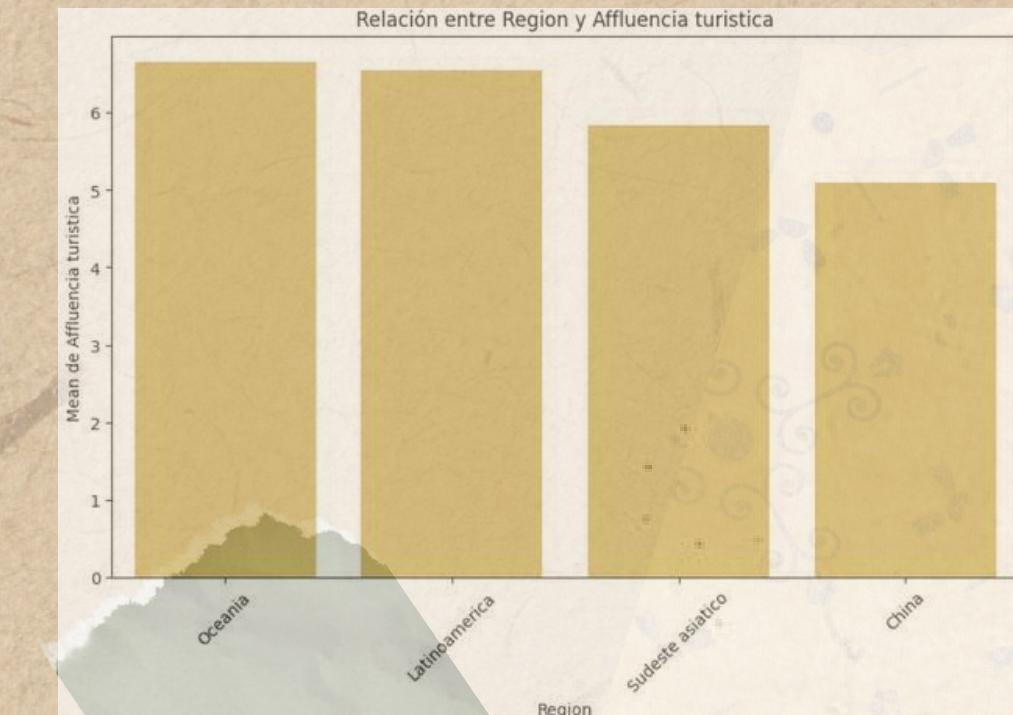
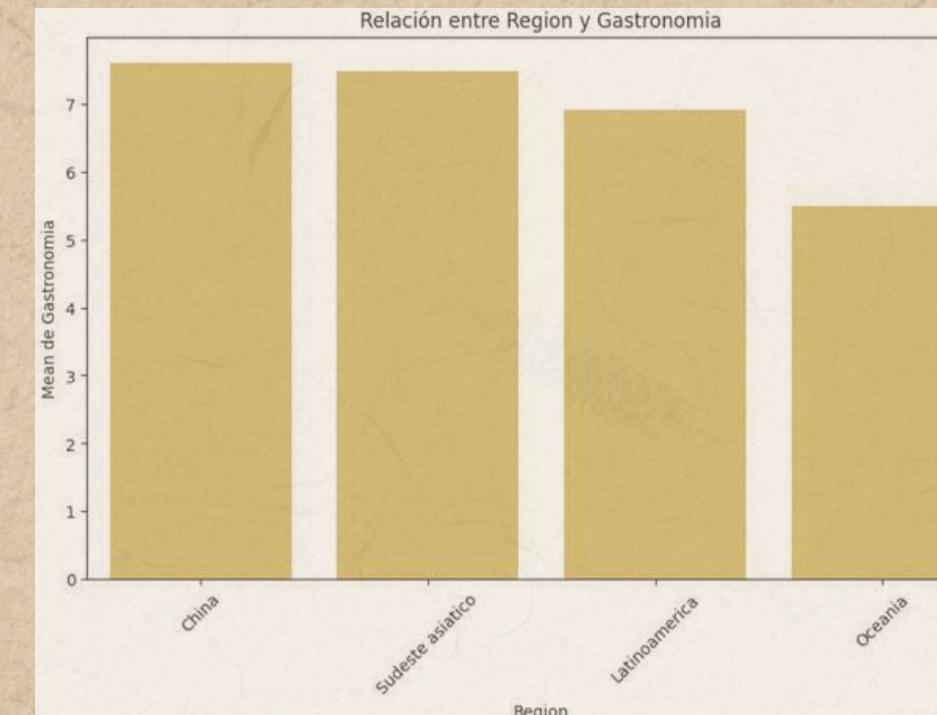
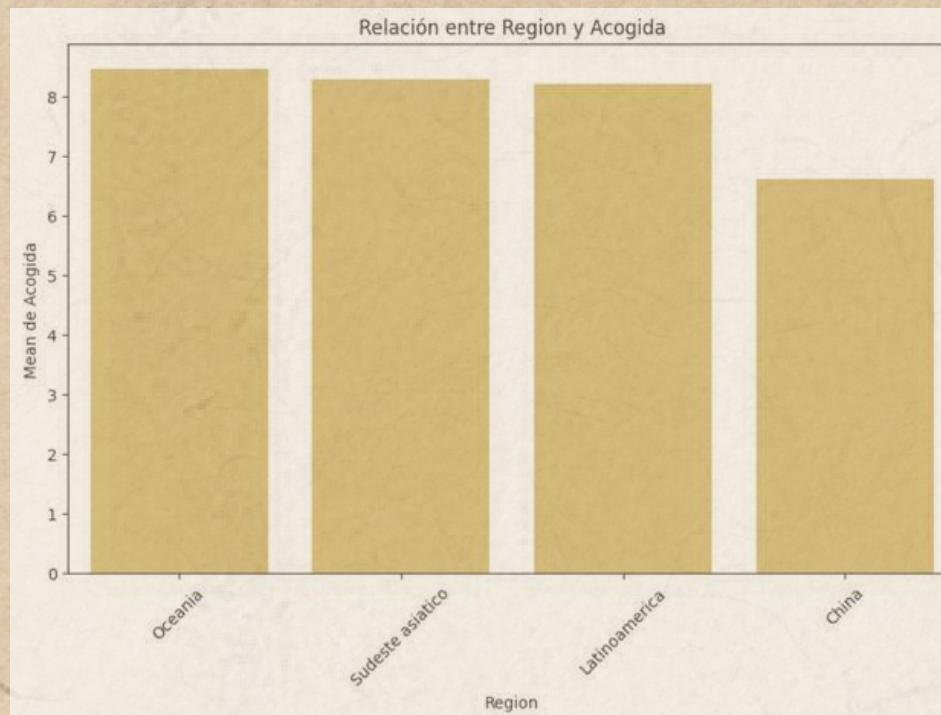
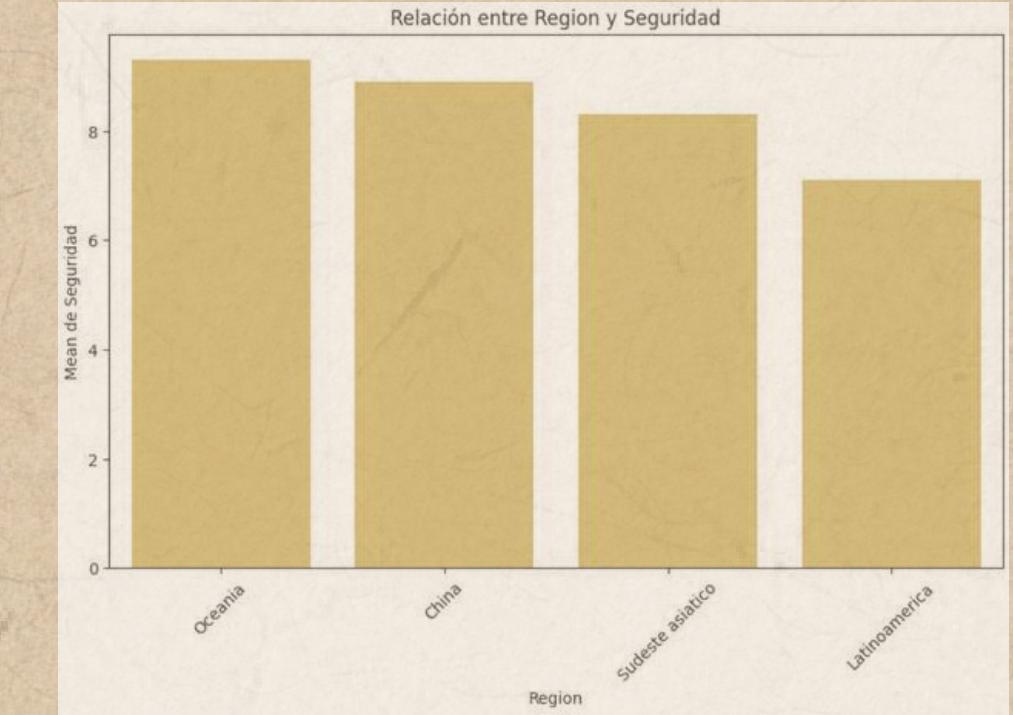
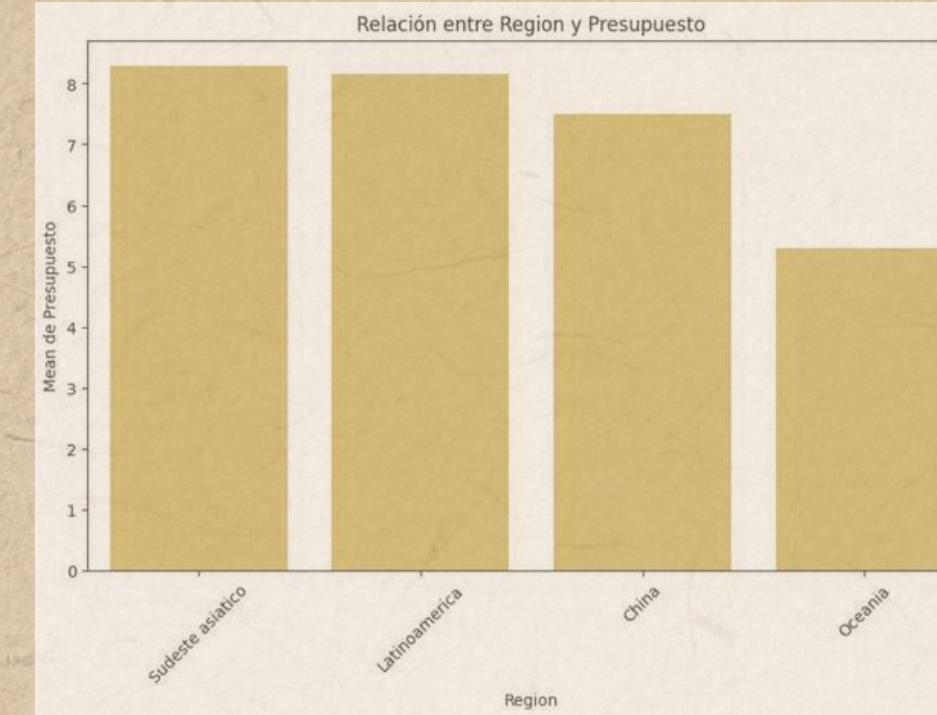
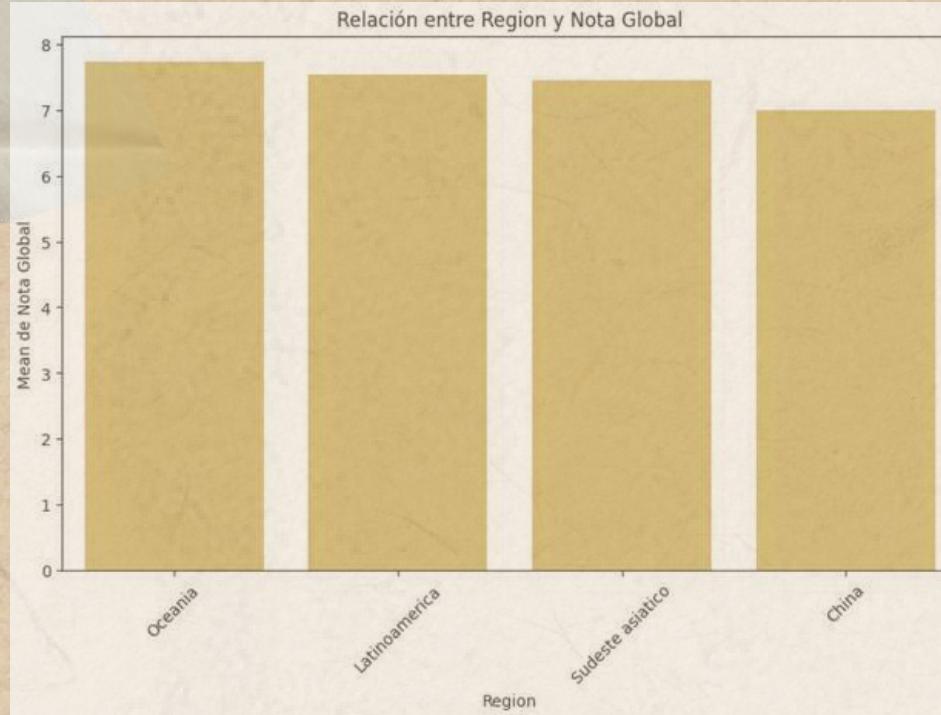
Turismo y alojamiento (7 categorías)

Nivel de Inglés (2 categorías)

Transporte (4 categorías)

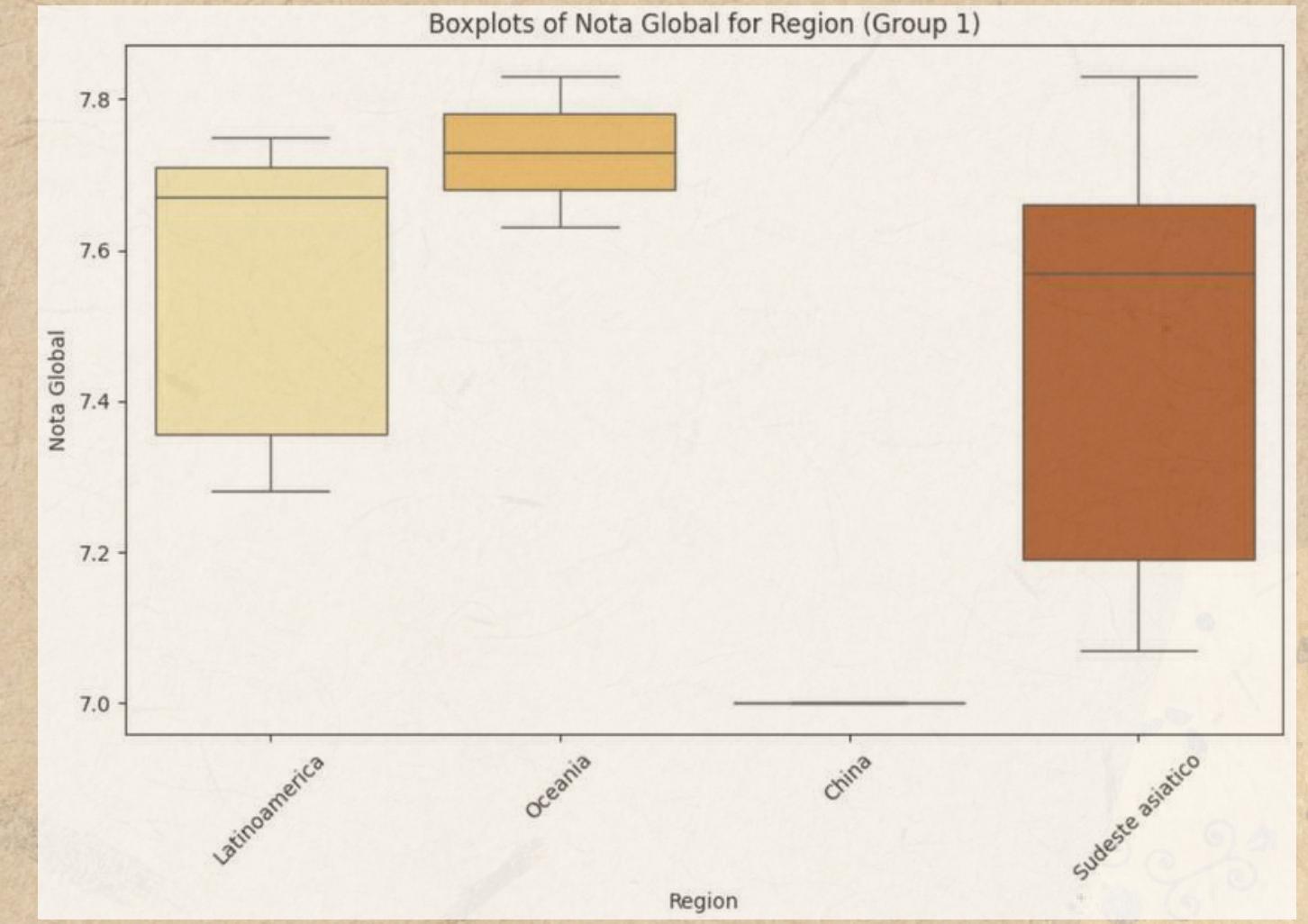
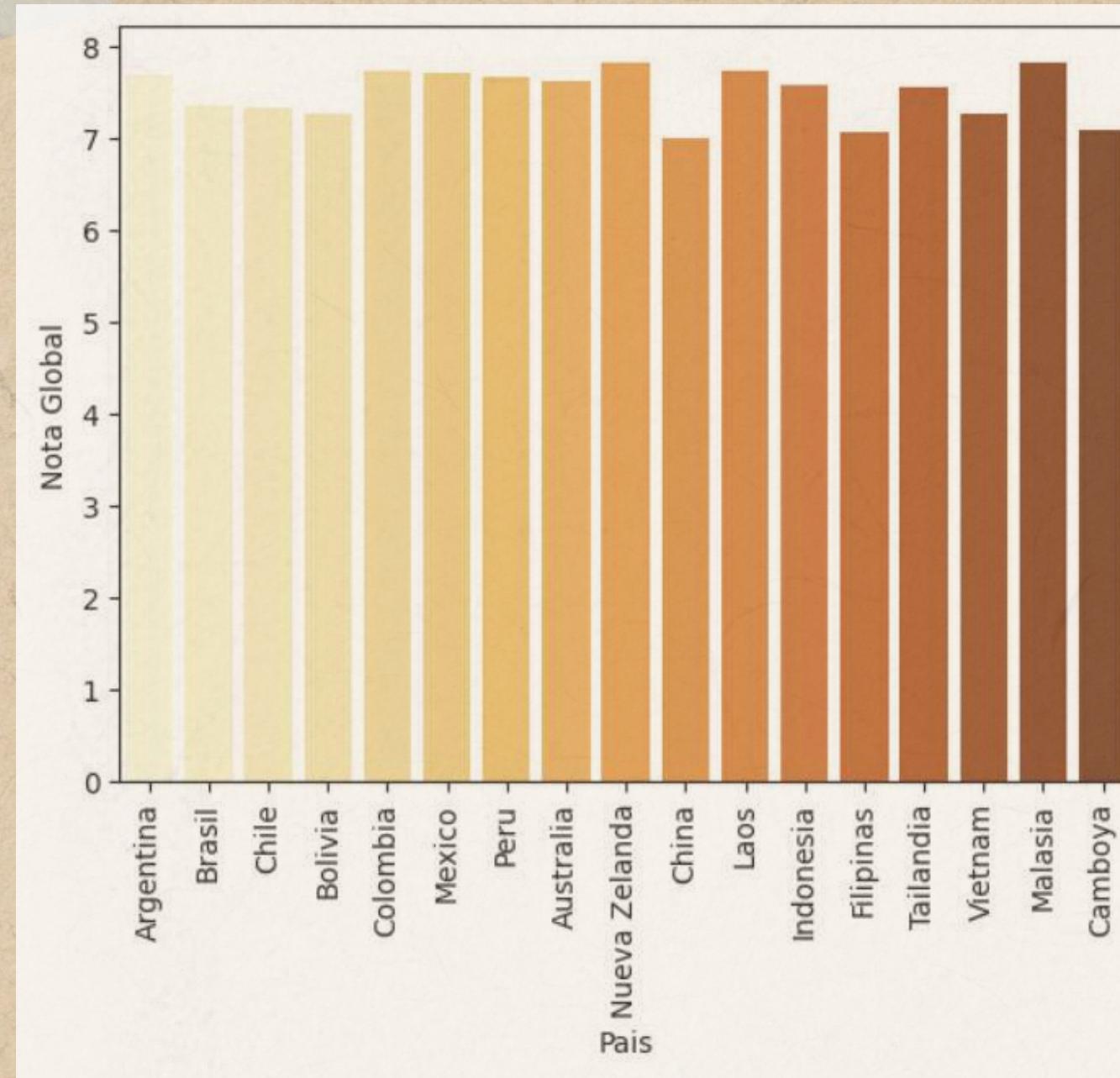
# Analisis Univariante

Todas las variables son Numérica continua



# Analisis Univariante

Todas las variables son Numérica continua



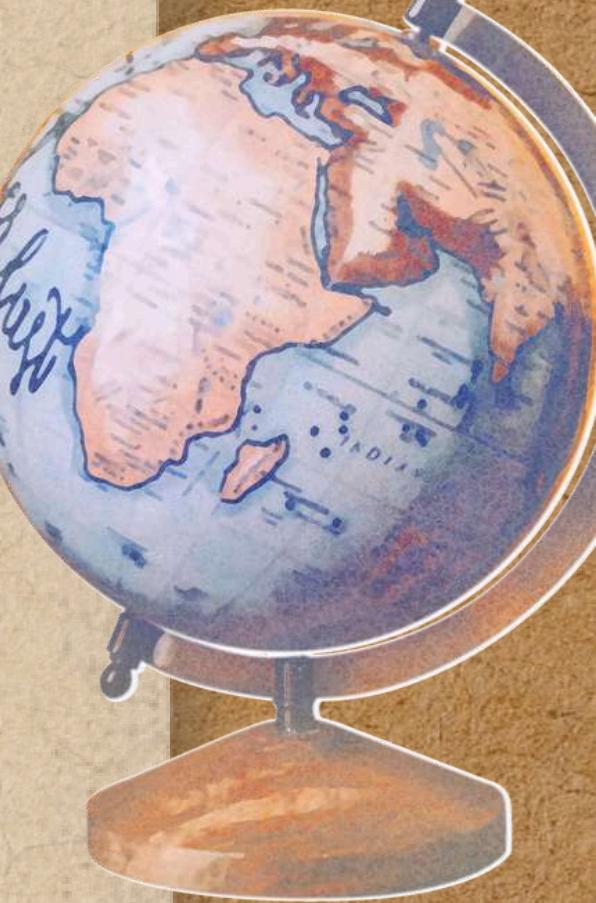
# Analisis Univariante

En general, respecto a la nota global, Oceanía resulta más atractiva para los turistas, con Latinoamérica (LA) y el Sudeste Asiático (SA) bastante cerca, y China en último lugar.

Como era de esperar, China supera por mucho el PIB de las otras tres regiones. El Sudeste Asiático ocupa la última posición con un valor de  $4.96e+05$ .

El Sudeste Asiático parece estar mas preparado para el turismo siguiendo parametros de personas que hablan ingles y % de población que trabaja en turismo.

Los países mas economicos son tambien los mas inseguros y sucios pero tambien ofrecen una mejor acogida en general al turista.



# Ejemplo: Análisis multivariable (análisis gastronomía)

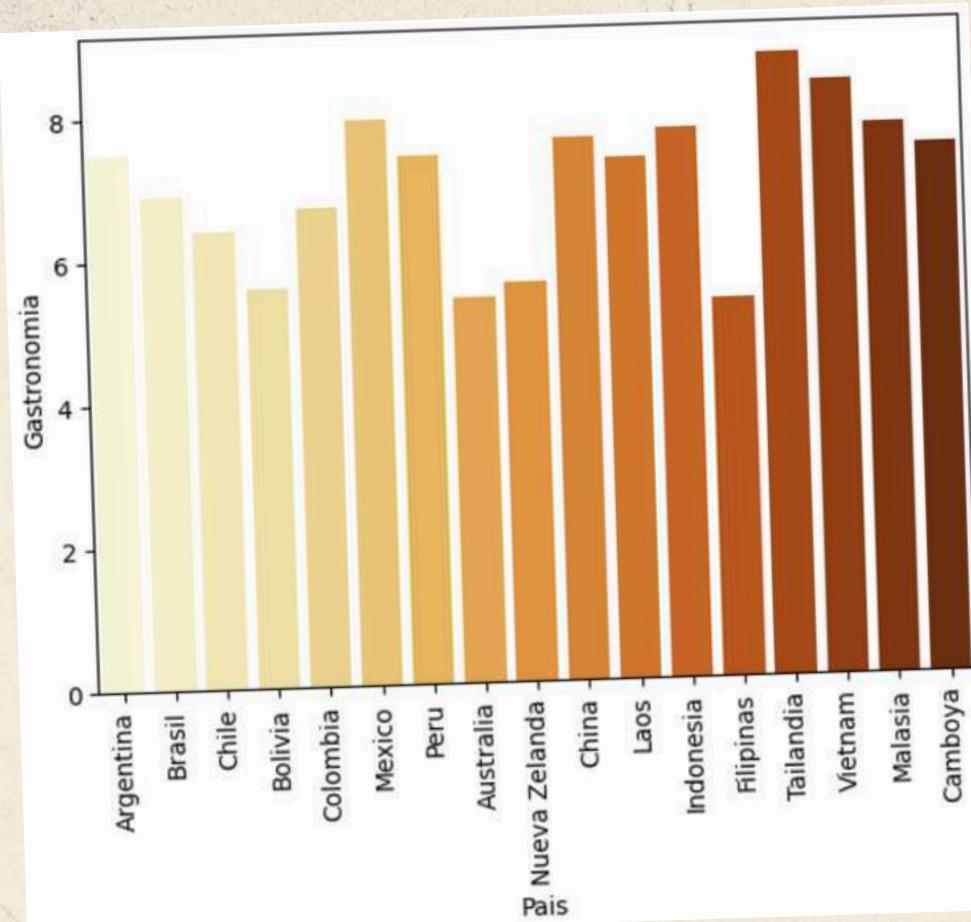
## tabla base



	Pais	Gastronomia	PIB (U.S. dollars)	top 50 restaurantes	Nota gastronomica	Ranking Gastronomico	Platos en top 100
0	Argentina	7.5	683533.0	7	4.28	25.0	2
1	Brasil	6.9	2125958.0	21	4.39	16.0	0
2	Chile	6.4	343823.0	8	4.24	32.0	1
3	Bolivia	5.6	56339.0	0	3.88	97.0	0
4	Colombia	6.7	427766.0	5	4.34	18.0	1
5	Mexico	7.9	1692640.0	27	4.52	3.0	3
6	Peru	7.4	303293.0	29	4.40	14.0	2
7	Australia	5.4	1742461.0	32	4.09	53.0	0
8	Nueva Zelanda	5.6	253436.0	0	3.96	82.0	0
9	China	7.6	19231705.0	7	4.45	10.0	8
10	Laos	7.3	16322.0	0	3.88	99.0	0
11	Indonesia	7.7	1429743.0	0	4.48	7.0	5
12	Filipinas	5.3	497495.0	0	4.21	36.0	0
13	Tailandia	8.7	546224.0	19	4.26	28.0	0
14	Vietnam	8.3	490970.0	0	4.34	19.0	1
15	Malasia	7.7	444984.0	0	4.20	37.0	0
16	Camboya	7.4	49799.0	0	3.00	125.0	0

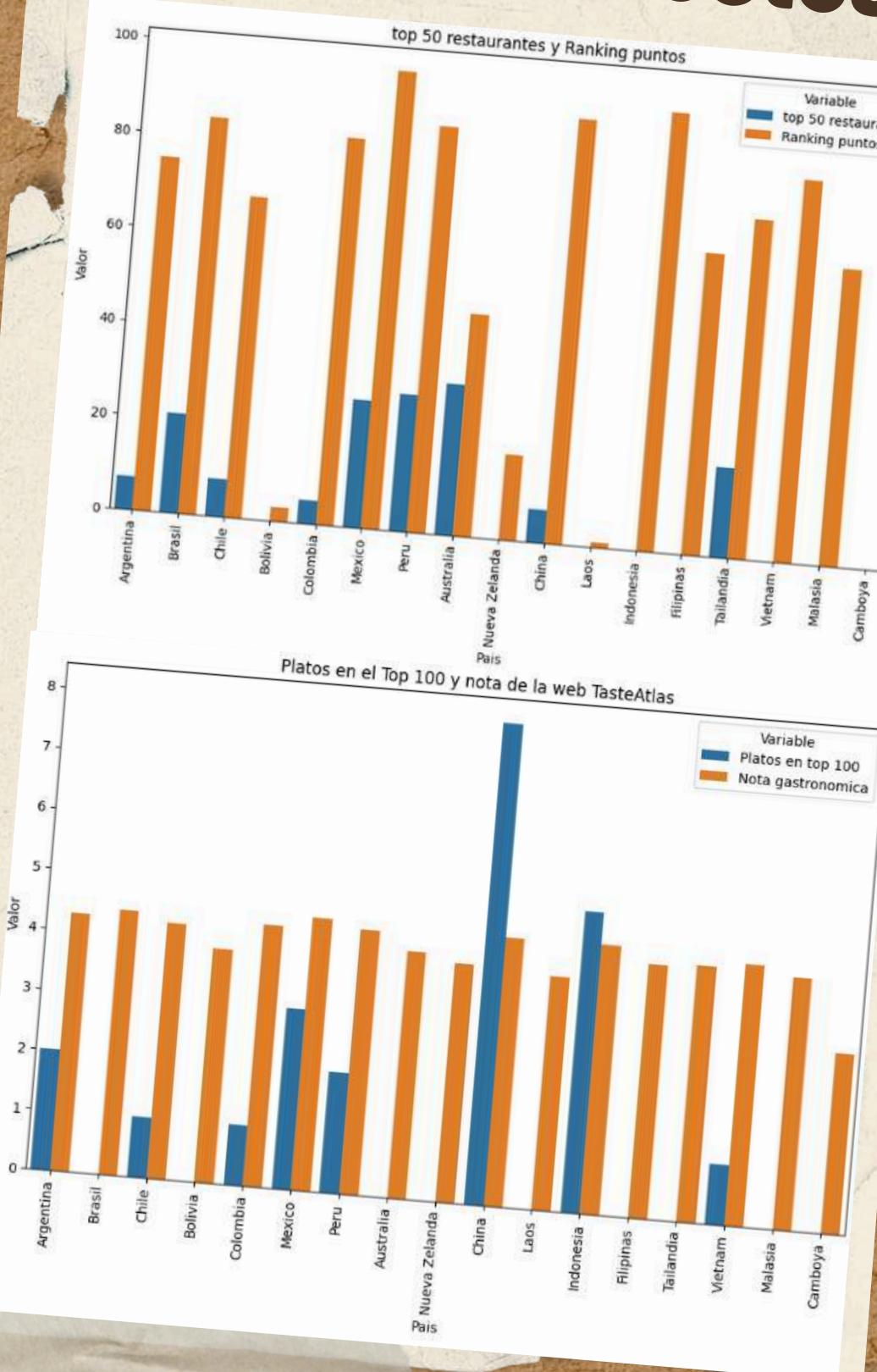
# Ejemplo: Análisis multivariable (análisis gastronomía)

## Opiniones turistas



Vemos que los Tailandia, Vietnam y Colombia son los países mejor valorados por los turistas.

## Variables Reales



Aquí vemos las cuatro categorías de gastronomía. México, Indonesia y China son las mejores gastronomías según TasteAtlas.

En cuanto a restaurantes de calidad, Australia, Perú y México ocupan el top 3.

En la grafica de abajo vemos en azul la cantidad de platos de una gastronomía en el top 100. China sobresale con 8 seguido de Indonesia y México.

En cuanto a la nota gastronómica, los mismos tres países ocupan el top 3.

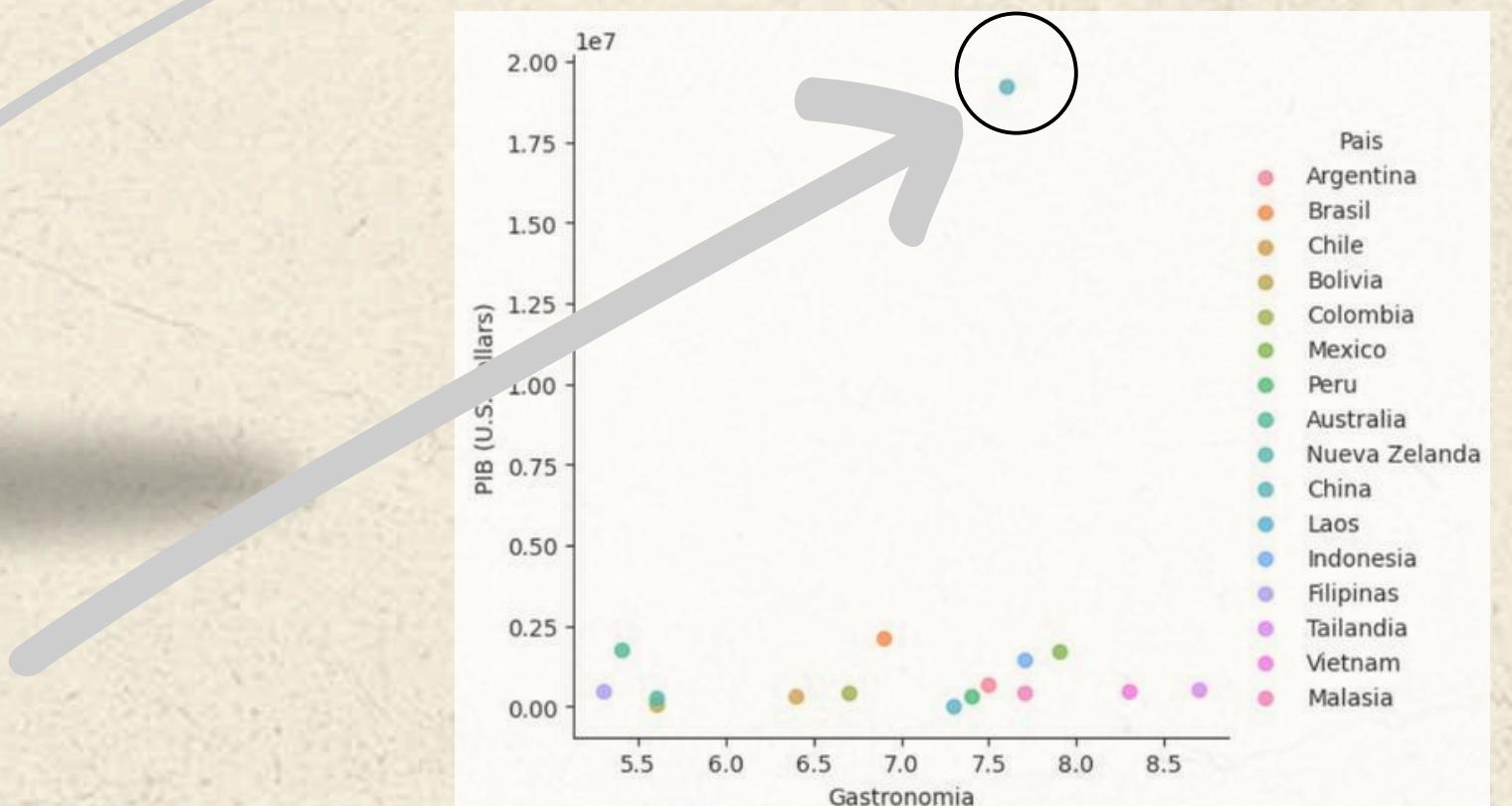
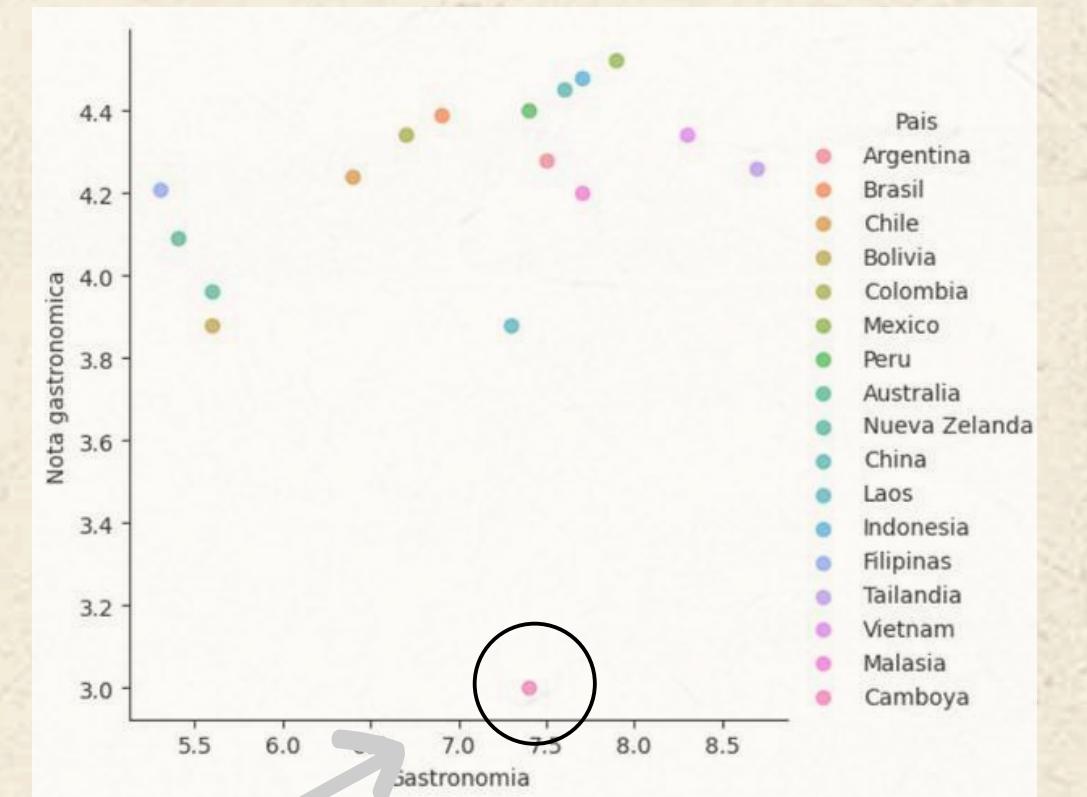
# Ejemplo: Análisis multivariable (análisis gastronomía)

## Eliminación de dos outlayers

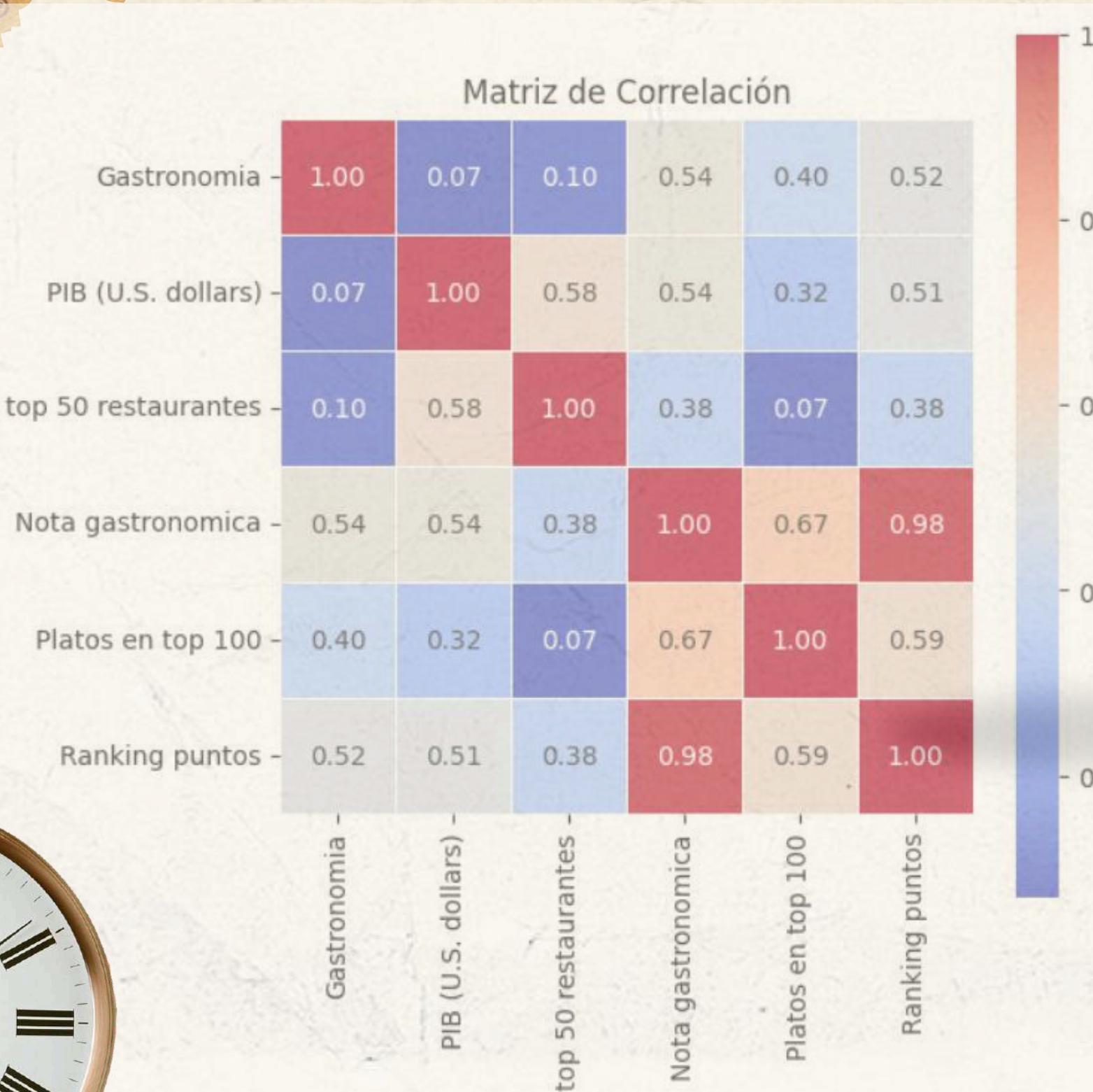
Vemos que Camboya parece ser un outlayer en términos de gastronomía.

Aunque conocemos la nota de los turistas (7.4) la nota gastronómica está trucada y no tiene ni restaurantes en el top 50 ni ranking gastronómico.

De igual manera, el anormalmente grande PIB de China nos dificulta la correlación de esta variable con el resto.



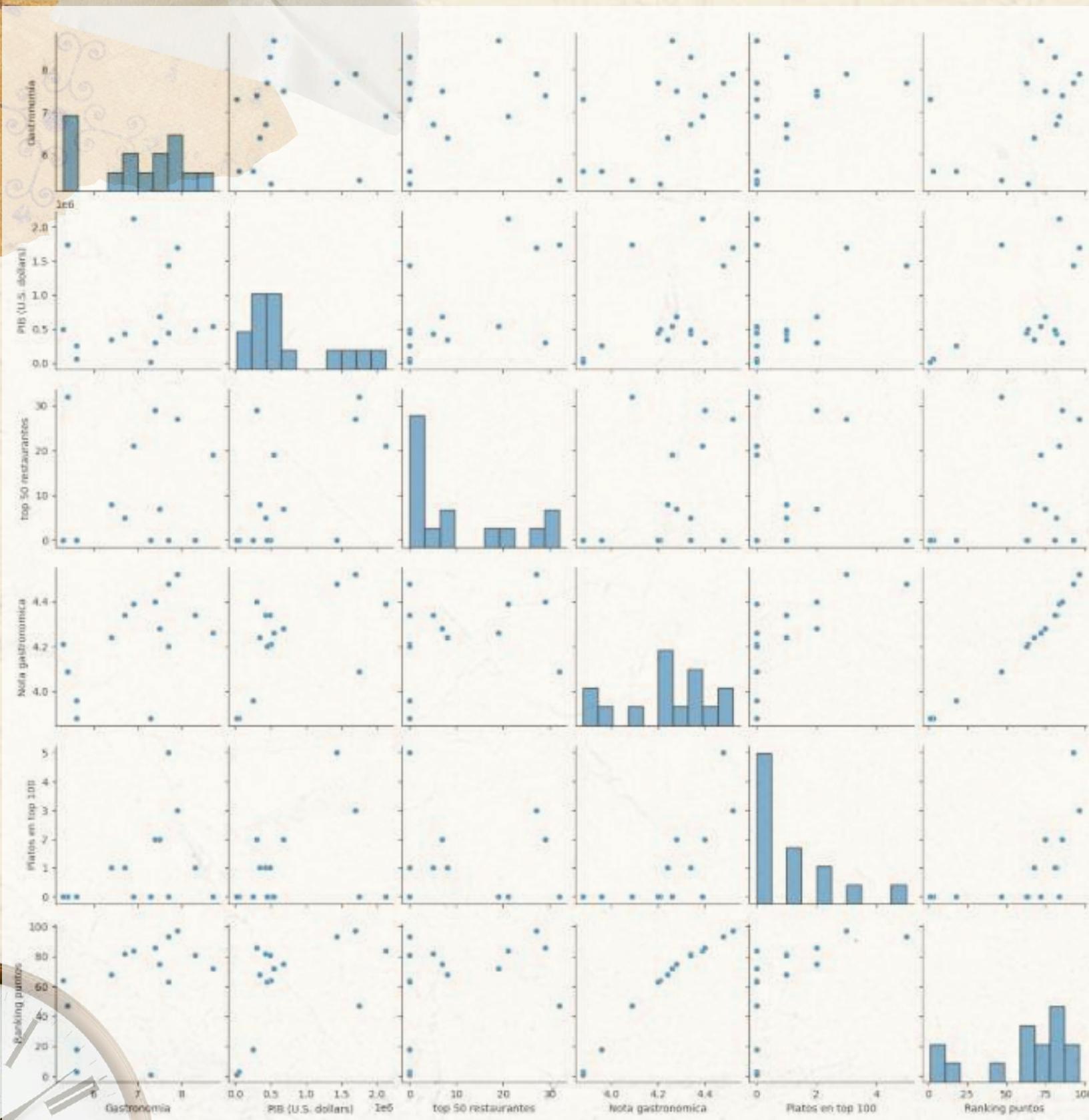
# Ejemplo: Análisis multivariable (análisis gastronomía)



Vemos que las correlaciones son mejores en general, con un 0.98 entre nota gastronómica y ranking y una mayor correlación entre numero de restaurantes en el top 50 y el PIB.

También la opinión de los turistas esta mas correlacionada con las notas y el ranking dados por atlasweb, aunque sigue sin haberla respecto a los restaurantes top y el PIB

# Ejemplo: Análisis multivariante (análisis gastronomía)



El gráfico pairplot no nos indica relaciones claras a la vista mas allá el ranking de gastronomía y la nota gastronómica. Observamos los p valor relacionando la opinión de los turistas y los demás valores.

Nota turistas vs Nota gastronómica

p valor =0.036

Nota turistas vs Top 50 restaurantes

p valor =0.73

Nota turistas vs Platos en Top 100

p valor =0.13

Nota turistas vs Ranking gastronómico

p valor =0.048



# Resumen resto análisis

## Presupuesto

Bolivia, Laos, Perú

Australia, N.Zelanda

Correlación



## Acogida

Colombia, Indonesia

Bolivia, China

Correlación



## Transporte

Malasia, Argentina

Camboya, Filipinas

Correlación



## Inglés

Australia, N.Zelanda

China, Vietnam

Correlación



## Hospedaje

Indonesia, Tailandia

Bolivia, Filipinas

Correlación



## Seguridad

Australia, N.Zelanda

Brasil, Colombia

Correlación



## Tranquilidad

Australia, N.Zelanda

China, Vietnam

Correlación



## Afluencia

Bolivia, Laos,

Malasia

China, Tailandia

Correlación



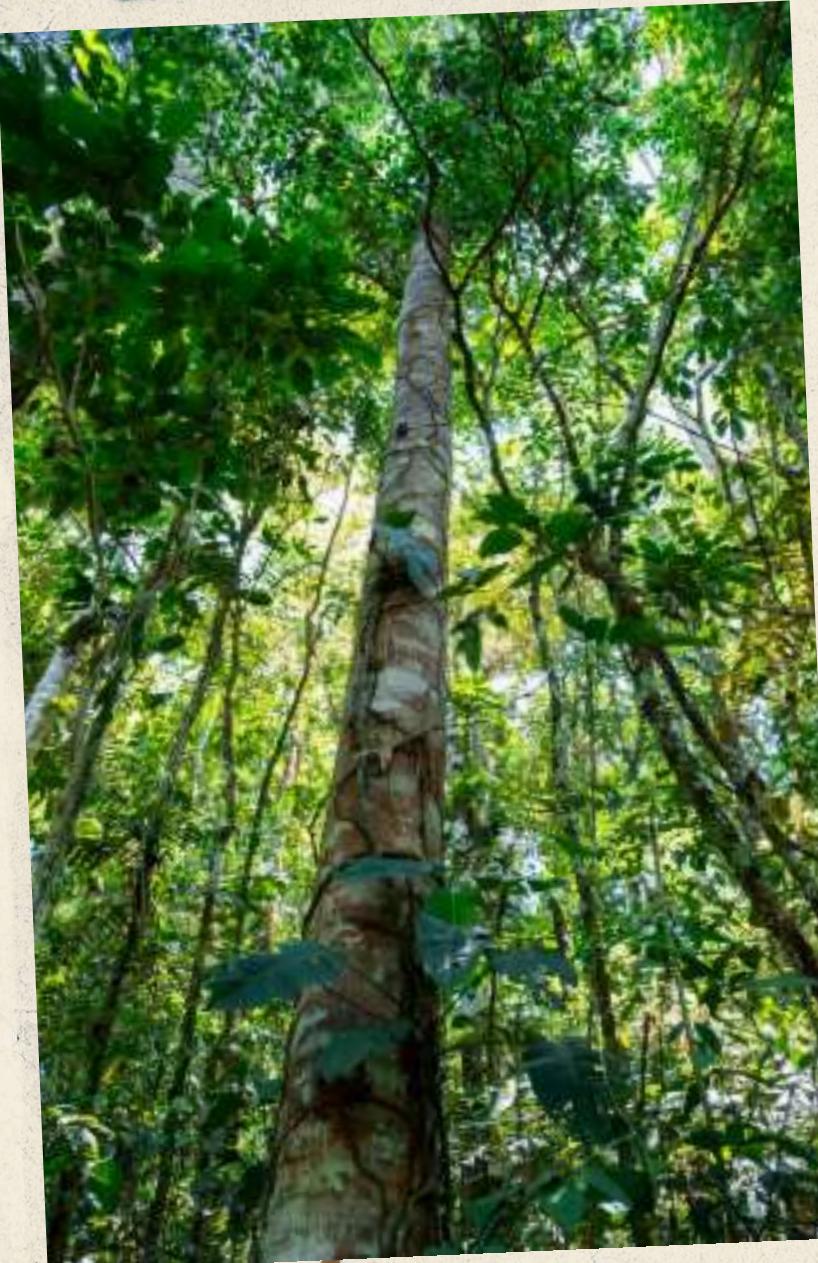
# Conclusiones

---

Parece que en general la opinión de los turistas coincide a “grosso modo” con variables realistas de los países y las tendencias generales cumplen lo previsto.

Existen casos como los hospedajes y el transporte que habría que analizar con mayor profundidad. Aunque posiblemente los servicios usados por los turistas son de mayor categoría que la media del país.

La totalidad de los análisis puede verse Online a través de Streamlit



# Caso práctico:

---

**Si fuera negocio,  
donde pondría mi  
hotel?**



# tabla base

Eliminamos países con alta inflación (Argentina, Bolivia, Laos)

	Pais	Presupuesto	Affluencia turistica	PIB (U.S. dollars)	Ahorro nacional bruto (% de PIB)	Inflacion anual %	Poblacion (Millones)	Inversion (% de PIB)	% paro
0	Argentina	6.5	6.7	683533.0	13.09	22.00	47.6	13.48	6.30
1	Brasil	8.4	6.3	2125958.0	14.47	4.82	212.8	16.76	7.15
2	Chile	6.5	6.6	343823.0	21.28	4.40	20.2	23.34	8.13
3	Bolivia	9.5	7.5	56339.0	11.39	15.53	12.6	14.95	5.05
4	Colombia	8.4	7.3	427766.0	16.34	5.10	53.1	18.59	10.00
5	Mexico	8.9	5.5	1692640.0	24.01	3.76	132.0	24.52	3.84
6	Peru	8.9	5.9	303293.0	19.61	1.36	34.3	17.92	6.45
7	Australia	5.3	6.5	1742461.0	23.60	2.10	27.2	23.98	3.70
8	Nueva Zelanda	5.3	6.8	253436.0	17.24	2.70	5.3	24.17	3.72
9	China	7.5	5.1	19231705.0	42.36	0.20	1416.0	40.47	5.10
10	Laos	9.1	7.6	16322.0	17.64	11.20	7.9	29.00	1.20
11	Indonesia	8.0	5.3	1429743.0	29.70	2.65	285.7	31.18	5.00
12	Filipinas	7.3	6.4	497495.0	19.21	1.70	116.8	22.64	4.48
13	Tailandia	8.2	3.9	546224.0	19.13	-0.72	71.6	17.95	1.00
14	Vietnam	8.5	5.3	490970.0	33.24	3.38	101.6	30.04	2.03
15	Malasia	8.5	7.1	444984.0	23.52	1.10	36.8	21.88	3.20
16	Camboya	8.4	5.2	49799.0	34.77	2.10	17.8	32.51	0.27

# tabla base

Eliminamos países con mayor cantidad de hoteles (Tailandia, Vietnam, Brasil)

	País	Affluencia turística	Hoteles	Número de turistas	Número de camas	Número de hoteles	Número de habitaciones	Ratio de ocupación
0	Argentina	6.7	7.3	7.399000e+06	711168.0	15713.0	266882.0	49.17
1	Brasil	6.3	7.4	6.353000e+06	2407892.0	15637.0	1011254.0	55.30
2	Chile	6.6	7.0	5.431000e+06	256806.0	5448.0	56106.0	40.45
3	Bolivia	7.5	6.4	3.233000e+05	51213.0	1453.0	31087.0	38.40
4	Colombia	7.3	7.7	1.396000e+06	606446.0	38523.0	361710.0	57.66
5	Mexico	5.5	7.7	5.112800e+07	1780426.0	26524.0	890213.0	53.38
6	Peru	5.9	7.3	1.119000e+06	532714.0	24819.0	308864.0	23.28
7	Australia	6.5	7.1	1.828000e+06	676638.0	6115.0	324471.0	66.50
8	Nueva Zelanda	6.8	7.3	9.960000e+05	289321.0	2397.0	78478.0	59.00
9	China	5.1	6.9	3.040200e+07	2046700.0	7337.0	1114100.0	38.35
10	Laos	7.6	7.2	8.864000e+05	189720.0	2403.0	99366.0	38.00
11	Indonesia	5.3	8.0	4.053000e+06	1124377.0	4129.0	747066.0	47.35
12	Filipinas	6.4	6.6	1.483000e+06	424746.0	6758.0	212373.0	30.40
13	Tailandia	3.9	7.9	3.991600e+07	1627449.0	20631.0	777391.0	29.34
14	Vietnam	5.3	7.3	3.837000e+06	1133940.0	38000.0	780000.0	37.00
15	Malasia	7.1	7.4	4.333000e+06	108602.0	5170.0	33491.0	28.20
16	Camboya	5.2	7.3	1.306000e+06	134706.0	3284.0	79239.0	36.26

inflación



# tabla base

Eliminamos países con mayores índices de criminalidad y de corrupción  
(Colombia, Chile, Perú, México y Camboya)

	País	PIB (U.S. dollars)	Seguridad	Tranquilidad	Índice de Criminalidad	Índice de Seguridad	Ranking Corrupción	Índice de Corrupción
0	Argentina	683533.0	7.6	8.0	63.32	36.68	99	63
1	Brasil	2125958.0	5.8	6.8	64.24	35.76	108	66
2	Chile	343823.0	7.6	8.0	60.54	39.46	32	37
3	Bolivia	56339.0	7.3	7.4	64.58	35.42	133	72
4	Colombia	427766.0	6.9	7.5	60.97	39.03	92	61
5	México	1692640.0	7.0	7.3	53.17	46.83	143	74
6	Perú	303293.0	7.6	7.5	66.72	33.28	128	69
7	Australia	1742461.0	9.1	8.7	47.41	52.59	10	23
8	Nueva Zelanda	253436.0	9.5	9.0	48.45	51.55	4	17
9	China	19231705.0	8.9	6.1	23.53	76.47	77	57
10	Laos	16322.0	8.9	8.8	44.24	55.76	117	67
11	Indonesia	1429743.0	8.4	7.6	46.06	53.94	101	63
12	Filipinas	497495.0	7.9	7.7	43.40	56.60	116	67
13	Tailandia	546224.0	8.5	7.3	36.84	63.16	112	66
14	Vietnam	490970.0	8.2	6.0	40.38	59.62	91	60
15	Malasia	444984.0	8.6	7.9	48.62	51.38	57	50
16	Camboya	49799.0	7.7	7.3	50.15	49.85	158	79

inflación ✓

Número Hoteles ✓

# tabla base

Eliminamos países con mayores costes de vida (Nueva Zelanda y Australia)

	Pais	Indice_Costo_Vida	Indice_comestibles	Indice_precios_restaurantes
22	Australia	62.97	58.31	143.79
27	Nueva Zelanda	59.35	56.68	136.35
76	Argentina	38.34	43.06	48.26
80	Mexico	37.64	37.39	51.12
85	Chile	35.78	34.27	56.74
89	Tailandia	35.35	22.02	51.90
97	Camboya	33.02	22.61	30.32
103	Malasia	31.12	22.14	84.08
107	Peru	29.97	24.16	47.93
109	Filipinas	29.45	18.47	40.50
110	China	29.43	19.93	88.08
113	Brasil	28.63	23.84	46.57
115	Colombia	28.00	23.05	40.98
123	Vietnam	25.91	15.28	45.98
126	Bolivia	25.50	20.73	45.56
130	Indonesia	24.59	14.44	33.29

inflación ✓

Número Hoteles ✓

Corrupción ✓

Criminalidad ✓

# tabla base

Eliminamos países con peor % de angloparlantes (China)

	Pais	PIB (U.S. dollars)	Total English speakers	% personas que hablan ingles
0	Argentina	683533.0	2.752681e+06	6.52
1	Brasil	2125958.0	1.200000e+07	5.56
2	Chile	343823.0	1.585027e+06	9.53
3	Bolivia	56339.0	2.876712e+06	4.38
4	Colombia	427766.0	2.012950e+06	4.22
5	Mexico	1692640.0	1.568626e+07	13.00
6	Peru	303293.0	2.250000e+05	1.24
7	Australia	1742461.0	2.316134e+07	91.10
8	Nueva Zelanda	253436.0	4.750056e+06	97.15
9	China	19231705.0	1.000000e+07	0.70
10	Laos	16322.0	2.400000e+06	4.45
11	Indonesia	1429743.0	8.596861e+07	30.80
12	Filipinas	497495.0	7.011794e+07	63.74
13	Tailandia	546224.0	1.712119e+07	27.16
14	Vietnam	490970.0	2.524224e+06	40.25
15	Malasia	444984.0	1.558000e+07	57.34
16	Camboya	49799.0	3.500000e+06	22.20

inflación ✓

Número Hoteles ✓

Corrupción ✓

Criminalidad ✓

Costo de vida ✓

# Si fuera negocio, donde pondría mi hotel?

Teniendo en cuenta los criterios de Inflación, Criminalidad, Corrupción, número de hoteles, coste de vida y nivel de inglés, las mejores opciones son:

**Malasia y Filipinas**



**Thank  
you!**

*NextGen Travel & Álvaro Martínez*

-----





México

China

# Bonus Track: y tu, ¿dónde irías?

descubrelo en Streamlit

Vietnam

Perú

Malasia

Argentina

Filipinas

