

SINCRONA

Machine Learning para el Modelamiento y Gestión de Sistemas Complejos

Sesión 7:





**Análisis preliminar de datos del mercado mayorista
de papa de Lima Metropolitana**

Eduardo Zegarra





Agenda

-  Algunas características del mercado de papa de Lima Metropolitana
-  La fuente y procesamiento inicial de datos: papa en MML
-  Tratamiento en Python
-  Análisis preliminar

1. Algunas características del mercado de papas de Lima Metropolitana



De la chacra a la mesa: los que se quedan con el mayor margen comercial de los alimentos

Hay siete etapas en el trayecto de un producto agrícola peruano desde el campo hasta el consumidor final. Gestión calculó la variación de precios en ese recorrido, para los cuatro productos más consumidos del país. ¿Qué intermediarios ganan más y por qué?



Whitney Miñán

whitney.minan@diariogestion.com.pe

Lima, 16/02/2023 07:21 a. m.

Desde Huánuco hasta Lima, una [papa](#) recorre más de 300 kilómetros. El tubérculo sale de la chacra y atraviesa hasta siete eslabones en la [cadena agrícola](#) antes de servirse en una mesa. En ese trayecto está la clave para entender cómo el precio del cultivo sube y quiénes se quedan con el mayor pedazo de esa venta. ¿Qué pasa en medio?

Christian Garay, director general de Estadística, Seguimiento y Evaluación de Políticas del Ministerio de Desarrollo Agrario y Riego ([Midagri](#)), explica a [Gestión](#) que los eslabones del proceso de comercialización agrícola son estos: el productor, el acopiador, el transportista, el comerciante mayorista, el distribuidor, el comerciante minorista y el consumidor final.

“La intermediación es necesaria, es parte del proceso productivo y no se puede evitar. Pero sí se puede, de alguna manera, tratar de reducirla para que ambos extremos de la cadena ganen en el mercado: mayor rentabilidad para el productor y precios accesibles para el consumidor”, refiere.

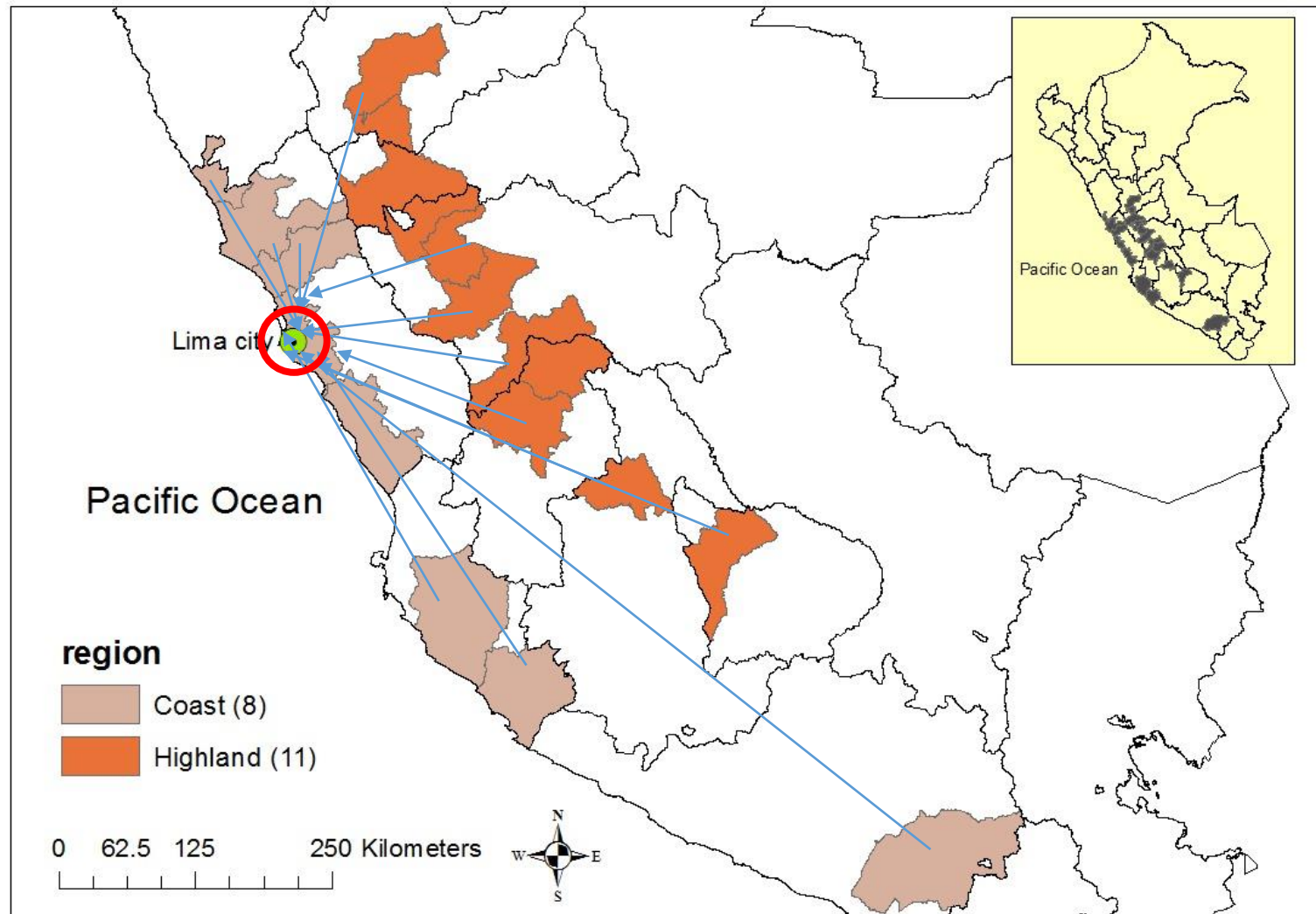
-Poder de negociación-

Otra forma de ver estas cifras es a lo largo del tiempo. Perú ha quedado envuelto en un contexto global de alza de los costos de producción, como los fertilizantes, y los precios finales también han subido. ¿Pero todos los eslabones de la cadena han podido elevar sus ganancias para mitigar el impacto de estos incrementos?

Aquí aparece otro concepto importante: el poder de negociación. Eduardo Zegarra, investigador principal de **Grupo de Análisis para el Desarrollo (Grade)**, dice que los precios del productor y los comerciantes mayoristas y minoristas han subido, *“lo que indica que estamos ante un proceso inflacionario generalizado, probablemente por una menor oferta y por el problema de los bloqueos”*.

Sin embargo, las magnitudes no son las mismas, porque no toda la cadena tiene el mismo poder para negociar. *“Se sabe por estudios que **en los últimos años aumentó mucho el poder que tienen los comerciantes y los acopiadores versus los productores, los cuales tienen limitado su poder de negociación, por su escala y su posición**”,* explica Zegarra.

Funcionamiento del mercado de papa de Lima Metropolitana



- Hay envíos de papa desde cada lugar (provincia) que se registran en el GMML diariamente
- Se registra la cantidad, origen y precio de cada embarque
- Se forma un solo precio promedio diario en el GMML para cada variedad.
- Usaremos la base de datos que se genera en este proceso (registro administrativo)

¿Cuáles son las principales variedades de papa en el Perú?...según ChatGPT



El Perú es un país de gran diversidad de papas, con más de 4,000 variedades de papas nativas identificadas. Algunas de las principales variedades de papa en Perú son:



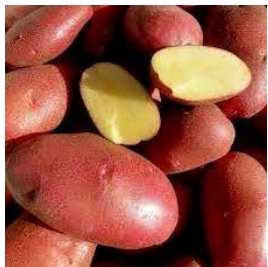
1. Papa Amarilla: también conocida como papa amarilla peruana, es una de las variedades más populares en el Perú. Tiene una textura suave y cremosa, y es ideal para preparar puré de papas, papas fritas y guisos.
2. Papa Huayro: también conocida como papa blanca, es una variedad de papa común en el Perú. Es un poco más firme que la papa amarilla, pero igualmente versátil en la cocina.
3. Papa Negra: como su nombre indica, esta variedad tiene una piel oscura y una pulpa amarilla. Es un poco más dulce que otras variedades de papa, y es ideal para hacer puré de papas, sopas y guisos.
4. Papa Peruanita: es una variedad pequeña y ovalada de papa, que se cultiva en las zonas andinas del Perú. Tiene una piel fina y una pulpa amarilla y firme, y es ideal para cocinar al horno o a la parrilla.
5. Papa Canchan: es una variedad de papa que se cultiva en la región de Cusco, en el Perú. Tiene una piel roja y una pulpa blanca y firme, y es ideal para cocinar al horno, asar a la parrilla o para hacer puré de papas.

Estas son solo algunas de las variedades de papas que se encuentran en el Perú, pero hay muchas otras, cada una con caracte

Regenerate response

La variedades de papa que ingresan al GMMML

Única



Blanca



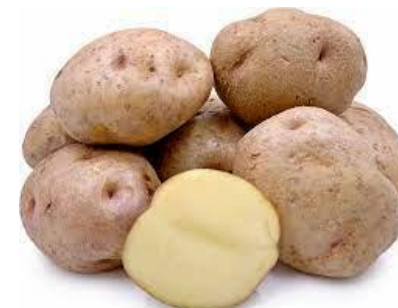
Yungay



Canchan



Perricholi



Papa "Color" ???

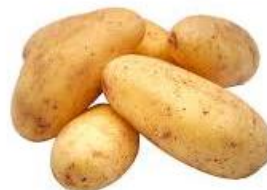
Huayro



Amarilla



Huamantanga



Peruanita



Negra Andina



Canales de comercialización de papa hacia Lima Metropolitana (Scott, 1985)

Figura III.5 Canales de comercialización de la papa en el Valle del Mantaro

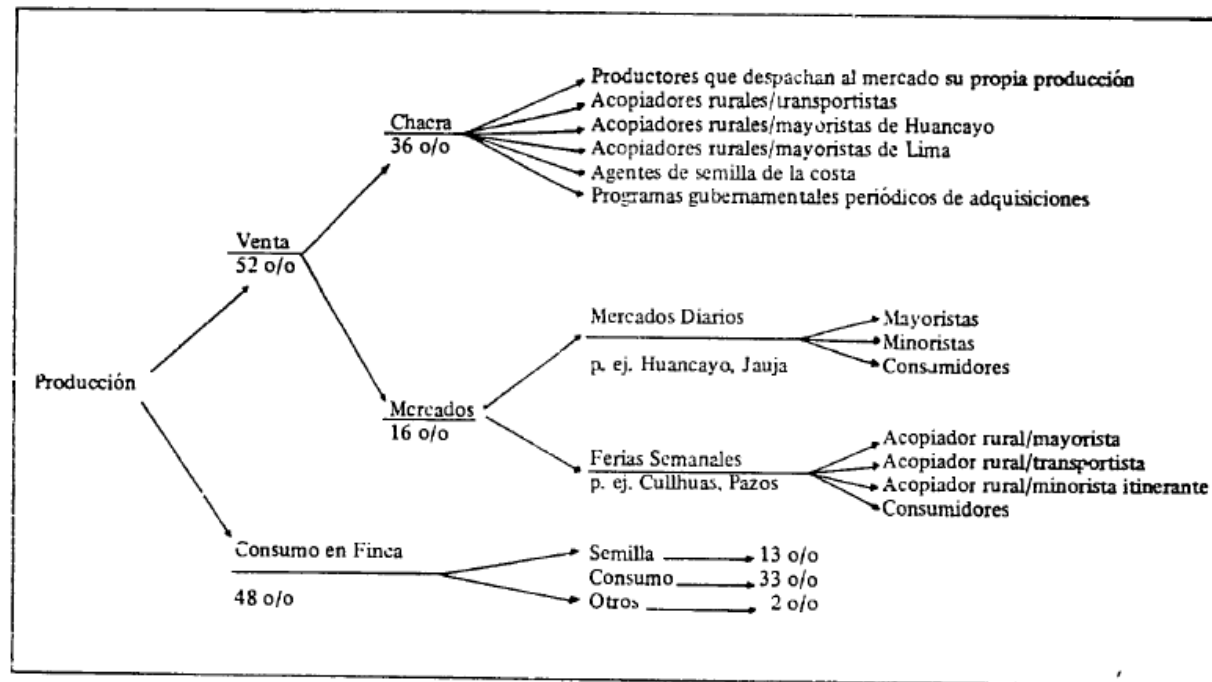
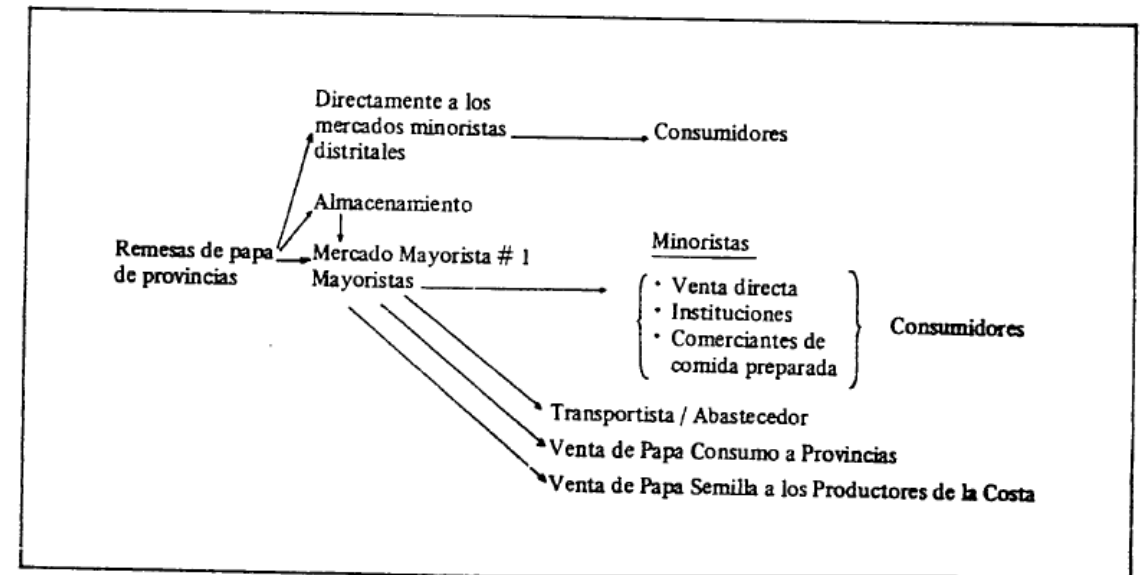


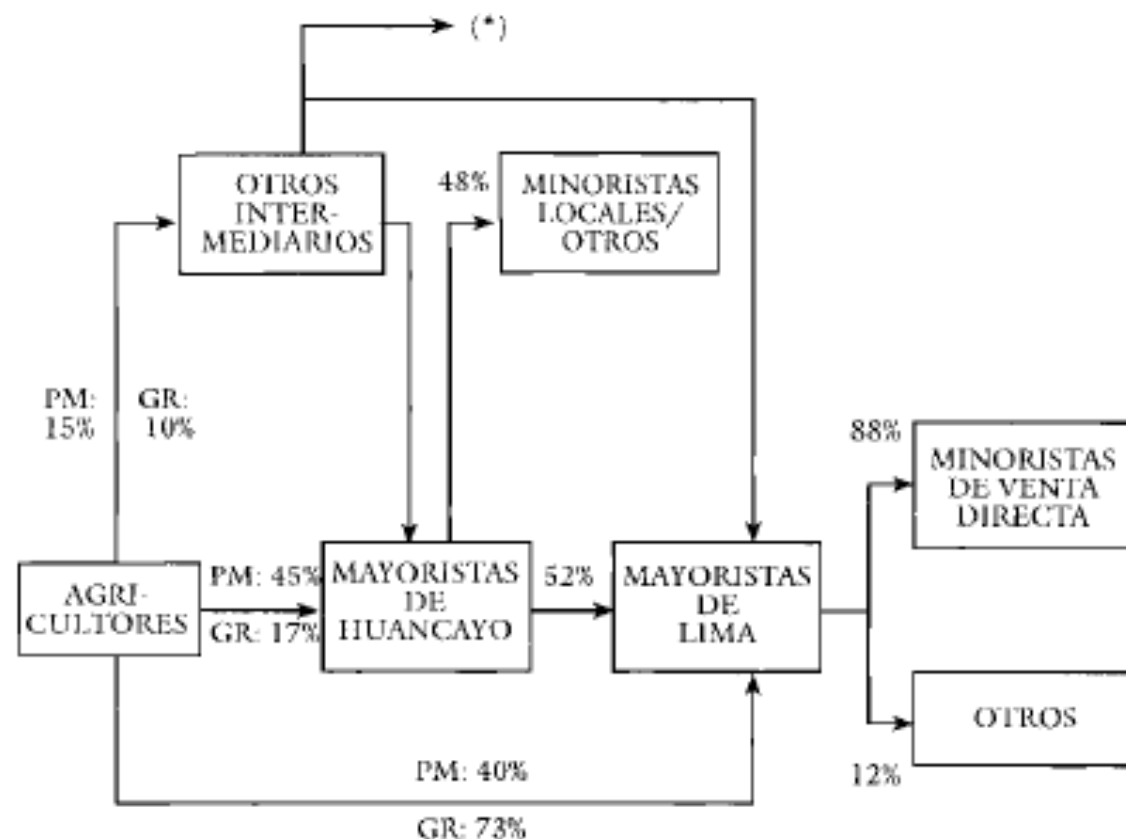
Figura V.1 Canales de comercialización de la papa en Lima



Estudio de Jorge Alarcón (1994)

Gráfico 2

Canales de comercialización de papa consumo del Valle del Mantaro a Lima:
distribución de las cantidades comercializadas



PM : Pequeños y medianos productores

GR : Grandes productores

Nota: Estos canales de mercadeo no incluyen las actividades de los pequeños agricultores de subsistencia.

Cuadro 3

Participación según tipo de comprador en el volumen negociado de papa para consumo, por tipo de productor (en porcentajes)

	Pequeños- medianos	Grandes	Total
- Acopiador mayorista-Huancayo	45.1	17.3	24.4
- Acopiador mayorista-Lima	39.8	73.2	65.5
- Otro acopiador intermediario	8.5	5.8	6.5
- Minoristas y consumidores vecinos	6.3	3.6	4.4

Elaborado en base a resultados de las encuestas a productores en el Valle del Mantaro, 1993.


2. Fuente y bases de datos



Fuente: SISAP—Ministerio de Agricultura

← → ↻ No es seguro | sistemas.midagri.gob.pe/sisap/portal2/mayorista/

Calendar Twitter / Cronología Pontificia Universid... GRADE_aula Vendor iadb TAI Diplomatura Superi...

 **PERÚ** Ministerio de Desarrollo Agrario y Riego

SISAP Versión 2.0
Información útil y oportuna

Consultar ▼ Herramientas | Inicio | ? Ayuda

11/08/2022 Día Intervalo de Tiempo Semanal Mensual Anual


Seleccionar Mercado... ▼

- ☐ Aceite +
- ☐ Aji fresco +
- ☐ Aji seco +
- ☐ Ajo +
- ☐ Alcachofa +
- ☐ Apio +
- ☐ Arroz +
- ☐ Arveja grano verde +
- ☐ Azucar comercial +
- ☐ Beterraga +
- ☐ Camote +
- ☐ Cebolla +

Variables:


- Precio Máximo
- Precio Promedio
- Precio Mínimo
- Volumen

SISTEMA DE ABASTECIMIENTO Y PRECIOS



Bienvenidos al Módulo de Mercados Mayoristas

Información útil y oportuna


PERÚ Ministerio de Desarrollo Agrario y Riego

SISAP Versión 2.0
 Información útil y oportuna

febrero 2019
 marzo 2020
 abril 2021
 mayo 2022

Consultar ▼ Herramientas | Inicio | Ayuda

Mensual Día Intervalo de Tiempo Semanal Anual

Seleccionar Mercado...

- ☐ Merlon
- ☐ Membrillo
- ☐ Naranja
- ☐ Olluco
- ☐ Otros agrícolas
- ☐ Otros cereales
- ☐ Otros frutícolas
- ☐ Pallar grano seco
- ☐ Pallar grano verde
- ☐ Palta
- ☒ Papa
- ☐ Papa seca

/variables:
 Precio Máximo
 Precio Promedio
 Precio Mínimo
Volumen

Procedencia:
 Lima
 Loreto
 Madre de dios
 Moquegua
 Pasco
 Piura
 Puno
 San martin
 Tacna
 Tumbes
 Ucayali
 Procedencia desconocida

SISTEMA DE AI
Y PR



Bier

M

N

M

Inform

SISAP

Versión 2.0

Información útil y oportuna

febrero 2019

marzo 2020

abril 2021

mayo 2022

Consultar

Herramientas

Inicio

Ayuda

Mensual

Día

Intervalo de Tiempo

Semanal

Anual

Seleccionar Mercado:

Otros cereales

Otros frutícolas

Pallar grano seco

Pallar grano verde

Palta

Papa

Papa seca

Papaya

Pepino

Pimiento fresco

Piña

Zanahoria

Variables:

Precio Máximo

Precio Promedio

Precio Mínimo

Volumen

Procedencia:

Lima

Loreto

Madre de dios

Moquegua

Pasco

Piura

Puno

San martin

Tacna

Tumbes

Ucayali

Procedencia desconocida

Lima Metropolitana: Volumen según producto y procedencia (sólo para volumen).

Periodos: Meses de enero a diciembre y Años de 2020 a 2021

Fecha	Papa Amarilla																					
	Volumen (t)																					
	Total	Tarma	Pasco	Jauja	Huanuco	Huancayo	Huamanga	Ambo	Andahuaylas	Tayacaja	Junin	Huanta	Huancavelica	Acobamba	Abancay	Cangallo	Daniel Alcides Carrion	Lima	Arequipa	Antabamba	Total	Tayacaja
DIC-2021	3265.00	619.00	133.00	486.00	255.00	702.00	41.00	3029.00	46876.00	34.00
NOV-2021	3294.00	326.00	44.00	110.00	...	400.00	...	2391.00	23.00	43599.00	22.00
OCT-2021	4669.00	618.00	150.00	141.00	...	755.00	...	2785.00	114.00	42.00	15.00	33.00	16.00	46490.00	43.00
SEP-2021	3164.00	291.00	82.00	381.00	...	500.00	...	1831.00	47.00	24.00	8.00	45035.00	84.00
AGO-2021	2996.00	258.00	160.00	117.00	93.00	322.00	135.00	1389.00	102.00	...	20.00	42069.00	285.00
JUL-2021	2318.00	193.00	51.00	58.00	...	237.00	274.00	1306.00	123.00	29.00	17.00	15.00	15.00	43395.00	221.00
JUN-2021	2774.00	464.00	54.00	64.00	...	373.00	419.00	1248.00	85.00	...	19.00	14.00	18.00	16.00	41257.00	277.00
MAY-2021	4215.00	823.00	261.00	136.00	...	524.00	626.00	1400.00	352.00	25.00	30.00	...	6.00	32.00	40372.00	871.00
ABR-2021	4952.00	497.00	320.00	116.00	173.00	498.00	456.00	2100.00	217.00	142.00	11.00	213.00	144.00	50.00	15.00	39204.00	626.00
MAR-2021	4345.00	453.00	181.00	35.00	...	491.00	162.00	2719.00	20.00	68.00	88.00	128.00	42091.00	204.00
FEB-2021	2352.00	293.00	42.00	185.00	91.00	1653.00	39.00	...	22.00	14.00	13.00	34756.00	139.00

<

Volumen de envíos mensuales a Lima desde cada provincia y por variedad de papa

- Para volumen se puede extraer mes y provincia de origen para cada año y variedad de papa
- Tomamos datos de volumen desde Enero 1997 a Mayo 2021 para todas las variedades

Precios: un solo precio mensual promedio por variedad (el precio que se forma en el mercado mayorista de Lima)

SISAP Versión 2.0
Información útil y oportuna

febrero 2019
marzo 2020
abril 2021
mayo 2022

Consultar

Herramientas Inicio Ayuda

Mensual Día Intervalo de Tiempo Semanal Anual

Seleccionar Mercado...

Otros cereales
Otros frutícolas
Pallar grano seco
Pallar grano verde
Palta
Papa
Papa amarilla
Papa blanca
Papa canchan
Papa color
Papa huamantanga
Papa huayro

variables:
Precio Máximo
Precio Promedio
Precio Mínimo
Volumen

Lima Metropolitana: Precio Promedio según producto.
Periodos: Meses de enero a diciembre y Años de 2020 a 2021

Fecha	Papa Amarilla Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Blanca Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Canchan Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Color Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Huamantanga Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Huayro Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Negra Andina Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Perricholi Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Peruanita Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Unica Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)	Papa Yungay Precio Promedio (\$/ x Kg., \$/ x Lt. o \$/ x Unid.)
DIC-2021	1.68	1.06	1.00	1.01	1.31	1.27	1.28	...	1.41	0.92	1.05
NOV-2021	1.40	1.08	0.99	0.87	1.24	1.15	1.28	...	1.14	0.79	1.09
OCT-2021	1.15	1.08	0.94	0.86	1.23	1.06	1.27	...	1.07	0.86	1.08
SEP-2021	1.48	1.13	1.13	1.13	1.47	1.16	1.24	...	1.10	1.11	1.13
AGO-2021	1.65	1.02	1.24	1.25	1.56	1.20	1.17	...	1.23	1.31	1.01
JUL-2021	1.47	0.92	1.37	1.44	1.58	1.07	1.27	...	1.25	1.48	0.92
JUN-2021	1.45	0.83	1.37	1.39	1.43	1.11	1.10	...	1.21	1.44	0.83
MAY-2021	1.67	0.78	1.29	1.29	1.31	0.96	0.96	...	1.30	1.30	0.78
ABR-2021	1.91	0.82	1.41	1.37	1.45	1.05	1.09	...	1.71	1.37	0.85
MAR-2021	2.24	0.77	1.17	1.29	1.42	1.34	0.97	...	2.02	1.13	0.78
FEB-2021	2.25	0.61	0.91	0.92	1.38	1.37	0.76	...	1.86	0.88	0.62

Datos de volumen en Excel

Lima Metropolitana: Volumen según producto y procedencia (sólo para volumen).													
Product Variable	Anual	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
1997													
Papa Amarilla													
Volumen (t)	930.96	97.47	132.7	107	219	60.12	39.69	66.96	147.2	...	33	26.58	1.2
Canete	27	13.2	13.8
Lima	4.2	4.2
Oxapampa	30.48	...	23.88	6.6
Pasco	846.39	93.27	108.8	87.24	215.9	60.12	37.44	53.88	129	...	33	26.58	1.2
Procedencia Sin Especificacion	6.09	2.25	3.84
Papa Blanca	16.8	3.12	9.24	4.44
Volumen (t)	94019	17901	5393	4407	4426	2781	3119	8759	7154	...	13731	16322	10026
Barranca	13194.1	3229.9	528.8	6.72	23.64	3.84	3.48	24.84	792.6	...	2878.3	3248.8	2453.1
Canete	24085.2	2507.9	244.8	109.7	12	33.59	1260	6950	3238	...	4221.1	4408.8	1098.5
Canta	48.12	24	12	1.56	...	3.36	3.6	3.6
Daniel Alcides	24	24
Huaral	21438	7360.7	1663	92.64	29.52	...	0.6	7.92	984.6	...	3500	4837.7	2961
Huachiriri	75.96	17.04	9.96	48.96
Huaura	7758.56	2494.8	388.6	24.84	...	2.16	...	70.08	226.4	...	1643	2013.7	894.98
Lima	6167.59	1523.4	628.6	98.76	23.64	45.6	36.72	133.4	766.5	...	905.41	1158.3	847.38
Oxapampa	226.21	10.2	28.2	35.64	20.4	76.6	...	31.04	24.13	...
Pasco	20294.7	733.92	1888	4005	4156	2607	1759	1499	1048	...	487.72	559.1	1552.7
Procedencia	270.89	16.44	10.44	25.56	93.68	35.25	25.14	19.68	8.76	...	5.52	22.2	8.22

Datos de precio en Excel

Lima Metropolitana: Precio Promedio según producto.													
Producto Variable	Anual	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
1997													
Papa Amarilla													
Precio Promedio Unid.)	0.95	0.9	0.87	0.8	0.71	0.78	1.01	1.1	1.2	1.09	0.92	1.05	1
Papa Blanca													
Precio Promedio Unid.)	0.64	0.56	0.47	0.4	0.46	0.61	0.75	0.74	0.78	0.84	0.78	0.68	0.65
Papa Canchan													
Precio Promedio Unid.)
Papa Color													
Precio Promedio Unid.)	0.66	0.55	0.49	0.44	0.48	0.61	0.75	0.74	0.79	0.85	0.8	0.71	0.67
Papa Huamantanga													
Precio Promedio Unid.)
Papa Huayro													
Precio Promedio Unid.)

- En el caso de los precios, se tiene un solo precio mensual por variedad (ya que es precio mayorista, no al productor)
- El precio es el promedio de los precios de cada día del mes, que a su vez es un promedio de precios de cada día

- Preprocesamos en Excel para tener datos en formato más legible desde Python, que usaremos para el procesamiento, limpieza y análisis exploratorio.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
producto	year	Variable	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_10	m_11	m_12
Papa Amarilla	1997	Abancay							16.80	22.20				
Papa Amarilla	1997	Ambo	97.80	30.00	26.40	12.00						10.56	10.60	51.90
Papa Amarilla	1997	Andahuaylas							89.40	87.00		96.12	20.16	
Papa Amarilla	1997	Arequipa		14.40										
Papa Amarilla	1997	Ascope	15.48											
Papa Amarilla	1997	Canete			13.20					13.80				
Papa Amarilla	1997	Concepcion					1.20	0.48				10.00		
Papa Amarilla	1997	Dos De Mayo								11.52		46.20	24.84	
Papa Amarilla	1997	Huamanga			23.64			18.00	0.36	18.00				
Papa Amarilla	1997	Huancavelica				3.60			2.40					
Papa Amarilla	1997	Huancayo	48.72	21.96	5.64	27.84	17.40	27.92	10.56	21.00		10.20	0.12	10.40
Papa Amarilla	1997	Huanuco	3351.72	2215.50	2455.56	2344.53	2329.57	1326.30	910.42	1409.92		2987.04	2427.74	1973.05
Papa Amarilla	1997	Huari							23.40					
Papa Amarilla	1997	Jauja			17.16	1.08								
Papa Amarilla	1997	Junin	15.60		20.40	4.80			4.80					
Papa Amarilla	1997	Lima	4.20											
Papa Amarilla	1997	Oxapampa		23.88	6.60									
Papa Amarilla	1997	Pachitea	383.04	170.52	145.32	58.08	14.04					167.32	144.62	70.42
Papa Amarilla	1997	Pasco	93.27	108.78	87.24	215.88	60.12	37.44	53.88	129.00		33.00	26.58	1.20
Papa Amarilla	1997	Paucartambo				6.60								
Papa Amarilla	1997	lencia Desconocida						2.25	3.84					
Papa Amarilla	1997	Sin Especificacion				3.12			9.24	4.44				
Papa Amarilla	1997	Tarma	7.92	28.80	12.00		27.24	8.52					7.20	
Papa Amarilla	1997	Tayacaja			22.80	26.40		0.84		20.00				
Papa Blanca	1997	Abancay			0.24				20.40	19.68				

Notar celdas sin información...

- Precios

rtapapeles Fuente Alineación Número Estilos Celdas Edición														
16														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	
producto	year	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_10	m_11	m_12	
Papa Amarilla	1997	0.90	0.87	0.80	0.71	0.78	1.01	1.10	1.20	1.09	0.92	1.05	1.00	
Papa Blanca	1997	0.56	0.47	0.40	0.46	0.61	0.75	0.74	0.78	0.84	0.78	0.68	0.65	
Papa Canchan	1997													
Papa Color	1997	0.55	0.49	0.44	0.48	0.61	0.75	0.74	0.79	0.85	0.80	0.71	0.67	
Papa Huamantanga	1997													
Papa Huayro	1997	0.82	0.75	0.69	0.53	0.67	0.79	0.89	0.99	1.01	1.05	1.09	0.85	
Papa Negra Andina	1997													
Papa Perricholi	1997													
Papa Peruanita	1997													
Papa Unica	1997													
Papa Yungay	1997													
Papa Amarilla	1998	1.00	1.02	1.43	1.52	1.43	1.59	1.67	1.72	1.67	1.45	1.07	1.10	
Papa Blanca	1998	0.62	0.58	0.78	0.79	0.93	1.08	1.42	1.54	1.05	0.66	0.53	0.48	
Papa Canchan	1998													
Papa Color	1998	0.63	0.57	0.77	0.79	1.00	1.11	1.45	1.53	1.05	0.72	0.60	0.51	
Papa Huamantanga	1998													
Papa Huayro	1998	0.78	0.71	0.96	1.02	0.98	1.12	1.25	1.58	1.55	1.65	1.25	0.82	
Papa Negra Andina	1998													
Papa Perricholi	1998													
Papa Peruanita	1998													
Papa Unica	1998													

Notar celdas sin información...

3. Procesamiento y limpieza con Python



Elementos básicos de Pandas (Python)

> I/O

Read and Write to CSV

```
>>> pd.read_csv('file.csv', header=None, nrows=5)
>>> df.to_csv('myDataFrame.csv')
```

Read and Write to Excel

```
>>> pd.read_excel('file.xlsx')
>>> df.to_excel('dir/myDataFrame.xlsx', sheet_name='Sheet1')
```

Read multiple sheets from the same file

```
>>> xls = pd.ExcelFile('file.xls')
>>> df = pd.read_excel(xls, 'Sheet1')
```

Summarize Data

`df['w'].value_counts()`

Count number of rows with each unique value of variable

`len(df)`

of rows in DataFrame.

`df.shape`

Tuple of # of rows, # of columns in DataFrame.

`df['w'].nunique()`

of distinct values in a column.

`df.describe()`

Basic descriptive and statistics for each column (or GroupBy).



pandas provides a large set of [summary functions](#) that operate on different kinds of pandas objects (DataFrame columns, Series, GroupBy, Expanding and Rolling (see below)) and produce single values for each of the groups. When applied to a DataFrame, the result is returned as a pandas Series for each column. Examples:

`sum()`

Sum values of each object.

`count()`

Count non-NA/null values of each object.

`median()`

Median value of each object.

`quantile([0.25, 0.75])`

Quantiles of each object.

`apply(function)`

Apply function to each object

`min()`

Minimum value in each object.

`max()`

Maximum value in each object.

`mean()`

Mean value of each object.

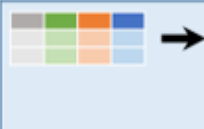
`var()`

Variance of each object.

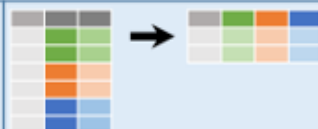
`std()`

Standard deviation of each object

Reshaping Data – Change layout, sorting, etc.



`pd.melt(df)`
Gather columns into rows.



`df.pivot(columns='var', values='val')`
Spread rows into columns.

Combine Data Sets

adf

x1	x2
A	1
B	2
C	3



bdf

x1	x3
A	T
B	F
D	T



Standard Joins

x1	x2	x3
A	1	T
B	2	F
C	3	NaN

`pd.merge(adf, bdf, how='left', on='x1')`
Join matching rows from bdf to adf.

x1	x2	x3
A	1.0	T
B	2.0	F
D	NaN	T

`pd.merge(adf, bdf, how='right', on='x1')`
Join matching rows from adf to bdf.

x1	x2	x3
A	1	T
B	2	F

`pd.merge(adf, bdf, how='inner', on='x1')`
Join data. Retain only rows in both sets.

x1	x2	x3
A	1	T
B	2	F
C	3	NaN
D	NaN	T

`pd.merge(adf, bdf, how='outer', on='x1')`
Join data. Retain all values, all rows.

Filtering Joins

Handling Missing Data

`df.dropna()`

Drop rows with any column having NA/null data.

`df.fillna(value)`

Replace all NA/null data with value.

Group Data



`df.groupby(by="col")`

Return a GroupBy object, grouped by values in column named "col".

`df.groupby(level="ind")`

Return a GroupBy object, grouped by values in index level named "ind".

All of the summary functions listed above can be applied to a group.

Additional GroupBy functions:

`size()`

Size of each group.

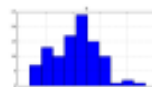
`agg(function)`

Aggregate group using function.

Plotting

`df.plot.hist()`

Histogram for each column



`df.plot.scatter(x='w', y='h')`

Scatter chart using pairs of points



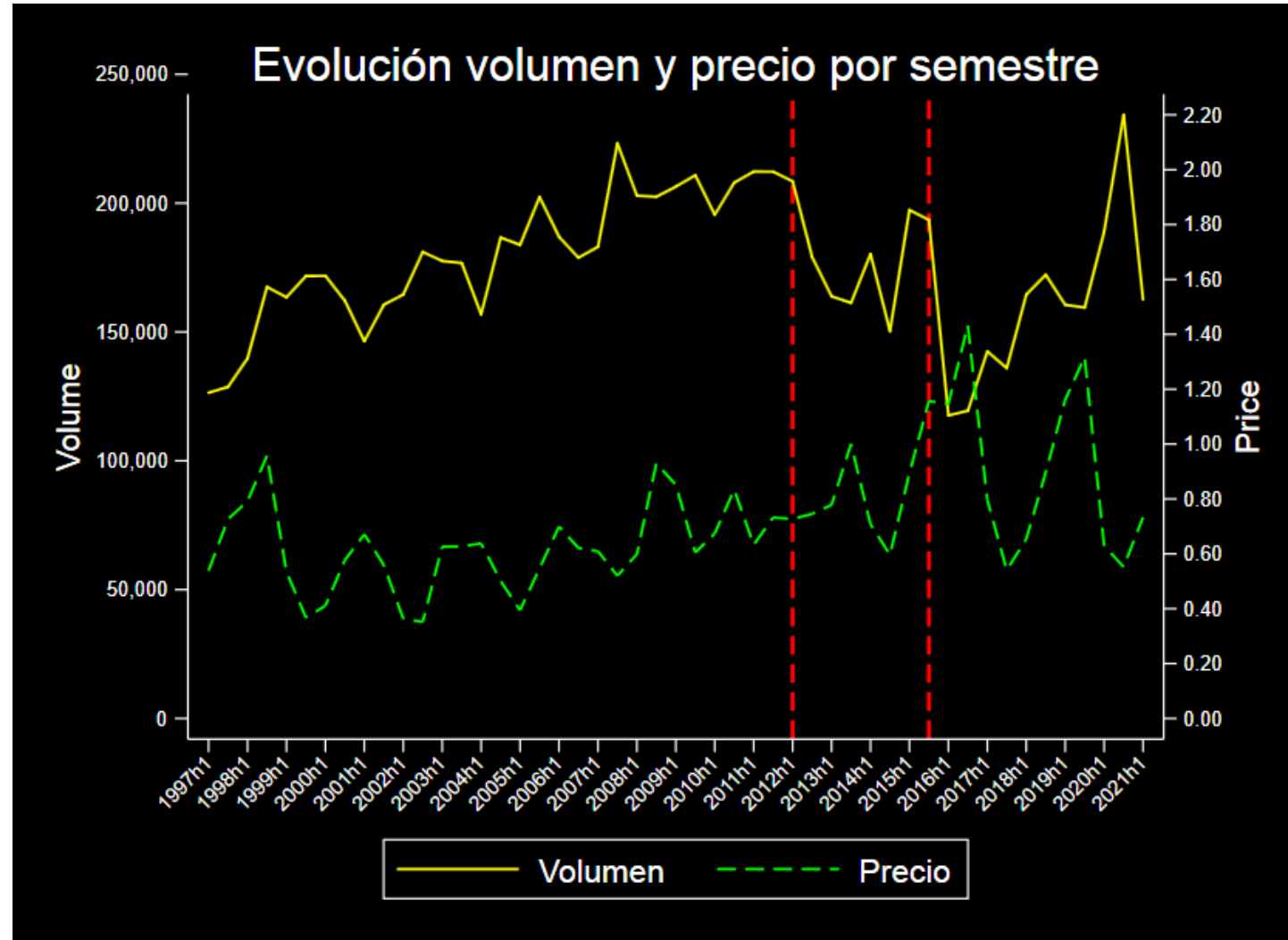
Cheatsheet for pandas (<http://pandas.py>)

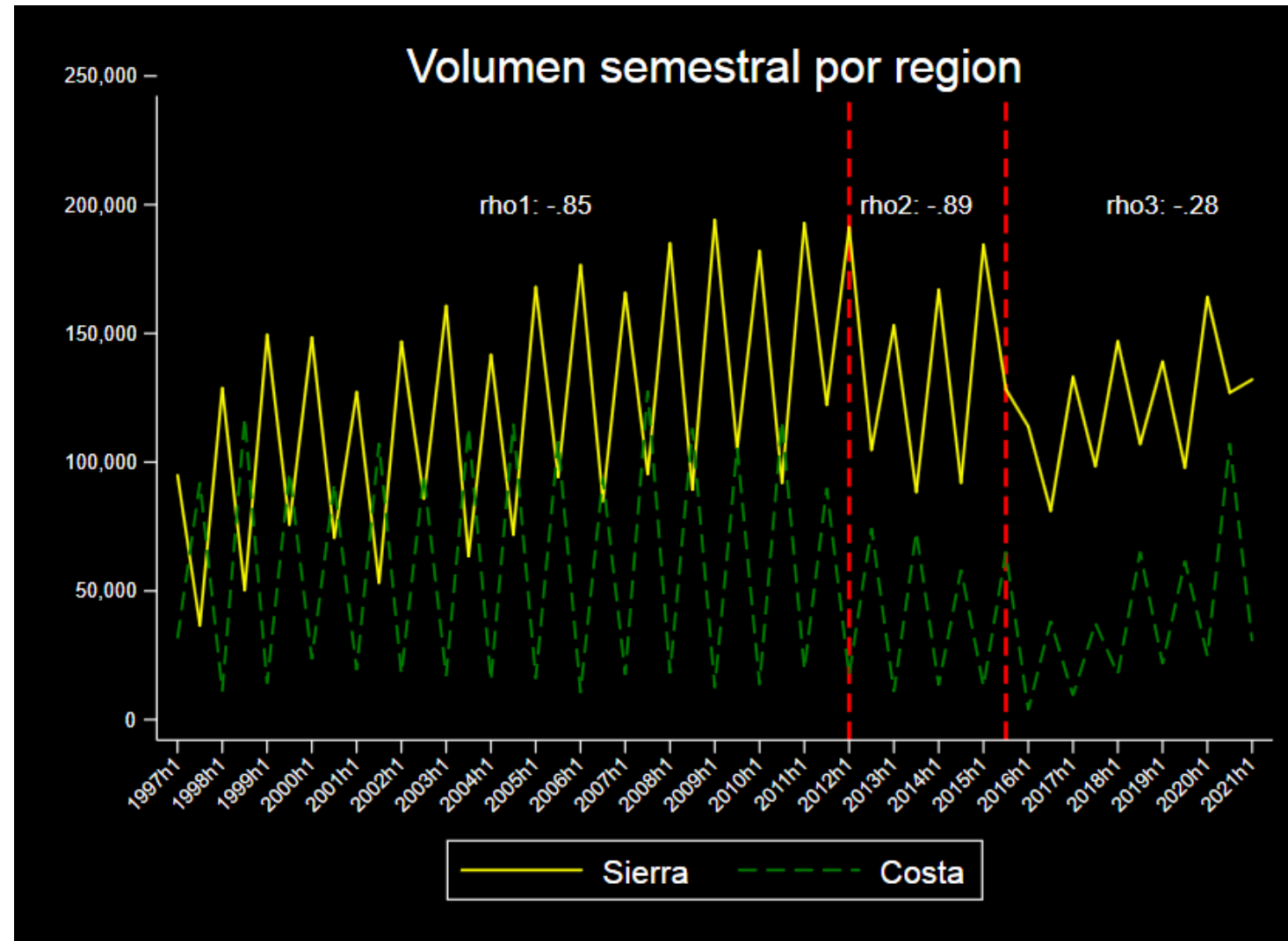
Trabajamos con el código que encontrarás en:

https://github.com/eduzegarra/grade_01/blob/main/proyecto_papa_v02.ipynb

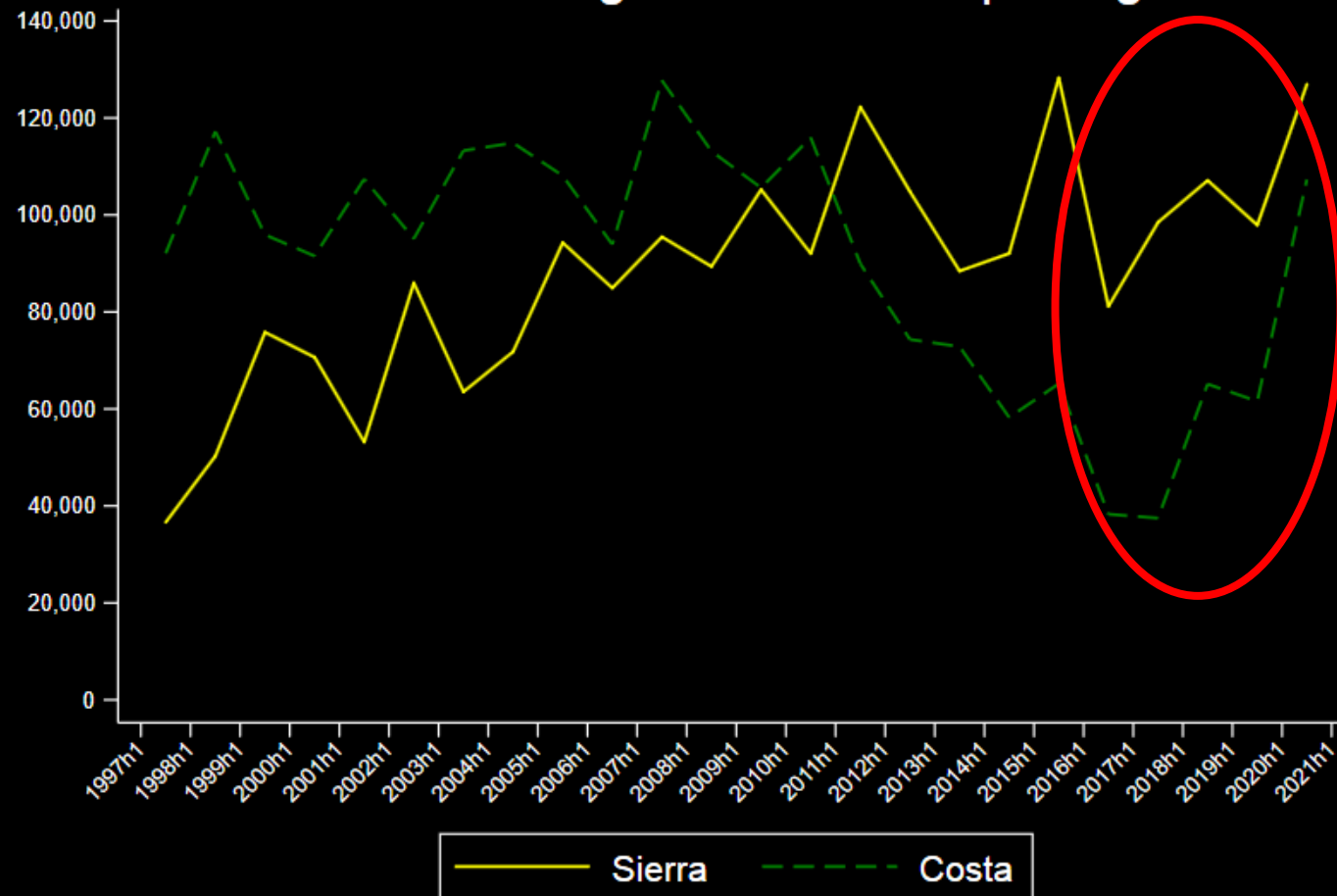
4. Análisis exploratorio







Volumen en segundo semestre por region



¡Muchas gracias!