



Universidad de Buenos Aires  
Laboratorio de Sistemas Embebidos  
Especialización en Inteligencia Artificial

Probabilidad y Estadística para la Inteligencia Artificial

Docente: Camilo Argoty

Grupo: 5

Fecha: 19-Jun-2024

- a1612 - Christopher Sebastian Charaf Herrera
- a1618 - Myrna Lorena Degano
- a1620 - Gustavo Julián Rivas

EXAMEN FINAL

Don Francisco es un pequeño comerciante de barrio con más de 40 años de experiencia en su negocio. Gracias a su esfuerzo, don Francisco cuenta hoy con dos supermercados de barrio, con los que sostiene su familia: El supermercado 'Santa Ana' y el supermercado 'La Floresta'.

Uno de sus hijos, Matías, quien recién inicia a cursar la Especialización en Inteligencia Artificial del LSE de la UBA, le propone hacer un análisis de las ventas durante el año anterior, con el fin de hacer pronósticos para el año siguiente, lo que a don Francisco le parece buena idea y le plantea algunas inquietudes:

1. Don Francisco necesita saber en qué mes se puede tomar unas vacaciones. También necesita saber en qué mes las ventas serán mejoras para embarcarse en una inversión importante. Por ello le pregunta a Matías, ¿cómo se comportan las ventas en cada mes?
2. De forma similar a la anterior, don Francisco necesita reacomodar los horarios de atención de sus supermercados a lo largo de la semana; también debe decidir en qué días debe contratar más empleados y en qué días podría contratar menos. Por ello pregunta, ¿cómo se comportan las ventas en cada día de la semana?
3. Don Francisco está un poco preocupado porque tiene la impresión de que la tienda 'La Floresta' vende menos que la tienda 'Santa Ana'.

Don Francisco le entrega a Matías el cuaderno donde tiene registrado el valor total de sus ventas en cada día del año. Con esta información, Matías construye sendas tablas por cada tienda en unas hojas de cálculo, en las cuales la primera columna corresponde a la fecha y la segunda corresponde al monto de las ventas, en dólares para evitarse dolores de cabeza con la inflación. Matías no se siente muy seguro de la tarea a realizar, así que les pide ayuda a ustedes para abordar el problema.

1. (3.5 puntos) Determinen funciones empíricas de distribución para las ventas en ambas tiendas durante cada uno de los meses del año. Determinen también aproximaciones a las funciones de densidad para las distribuciones anteriores. Determine intervalos de confianza empíricos para cada mes, para significancias del 95 % y el 99 %.
2. (3.5 puntos) Determine funciones empíricas de distribución para las ventas en ambas tiendas por cada día de la semana. Establezcan aproximaciones a las funciones de densidad para las distribuciones anteriores. Establezcan también intervalos de confianza empíricos para cada día de la semana, para significancias del 95 % y el 99 %.
3. (3 puntos) Realicen pruebas de hipótesis, con base en los datos, para aceptar o rechazar la intuición de Don Francisco, con significancias de 95 % y 99 %.

Deben realizar un informe en un documento pdf, en donde queden claros los siguientes aspectos:

1. Planteamiento de los problemas de investigación
  2. Procesamiento realizado a los datos
  3. Resultados encontrados
  4. Análisis y conclusiones
- Deben entregar también el código utilizado en el proceso.  
Muchos éxitos!!

## Informe - Grupo 5

Conjunto de datos .....	3
Procesamiento de los datos .....	3
Análisis de los datos .....	4
Por mes del año (La Floresta) .....	4
Por mes del año (Santa Ana) .....	8
Por día de la semana (La Floresta) .....	12
Por día de la semana (Santa Ana) .....	14
Comparación entre tiendas .....	16
Intervalos de confianza .....	17
La Floresta – Por mes del año .....	17
Santa Ana – Por mes del año .....	18
La Floresta – Por día de la semana .....	19
Santa Ana – Por día de la semana .....	20
Pruebas de Hipótesis .....	21
Conclusiones .....	22
Anexo .....	22

## Conjunto de datos

Registro de ventas diario para las tiendas “Santa Ana” y “La Floresta”.

	A	B	C	D
1	Fecha	Ventas		
2	1/1/2023	11391		
3	2/1/2023	16123		
4	3/1/2023	19609		
5	4/1/2023	16734		
6	5/1/2023	17984		
7	6/1/2023	17807		
8	7/1/2023	12287		
9	8/1/2023	13429		
10	9/1/2023	11088		
11	10/1/2023	14142		
12	11/1/2023	10070		

< >

Santa Ana

La Floresta

- Columna “Fecha”: formato fecha, con datos entre el 1/1/2023 y el 31/12/2023.
- Columna “Ventas”: formato numérico, incluyendo decimales, sin indicador de moneda (representa USD).

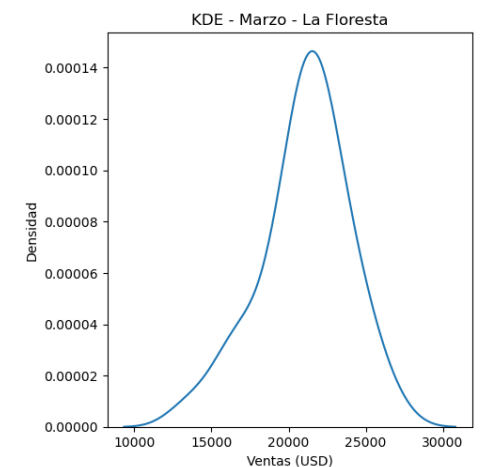
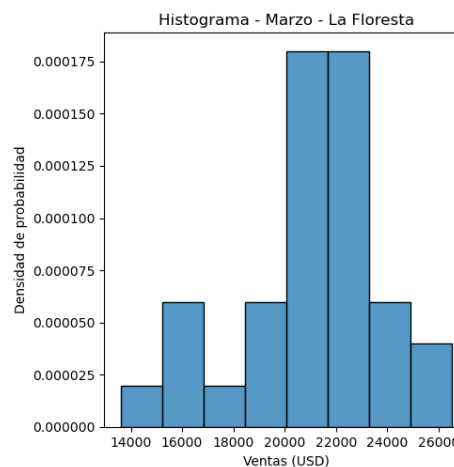
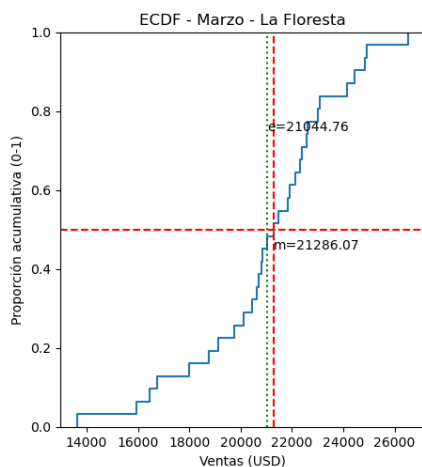
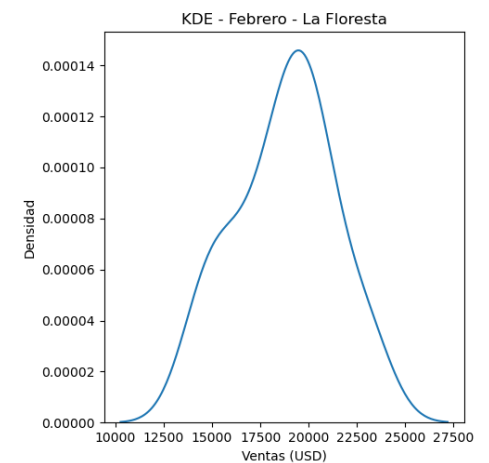
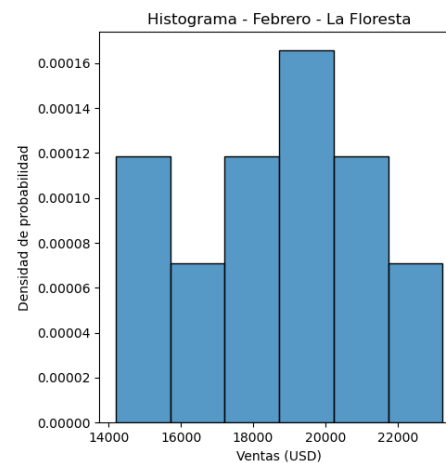
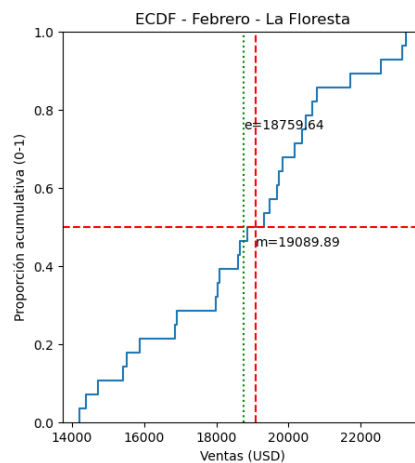
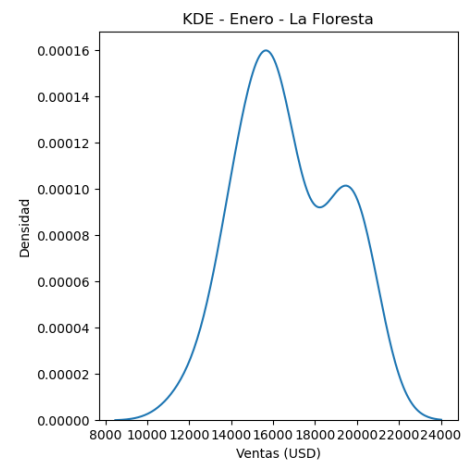
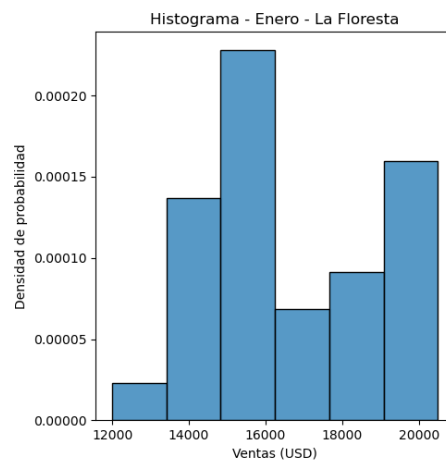
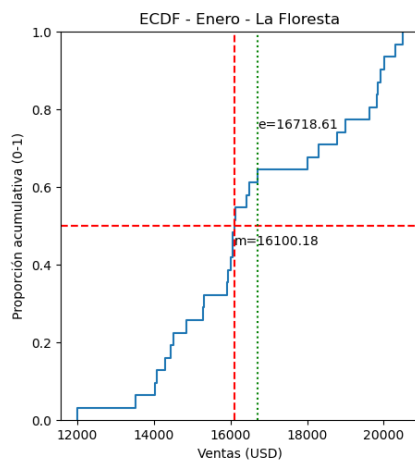
## Procesamiento de los datos

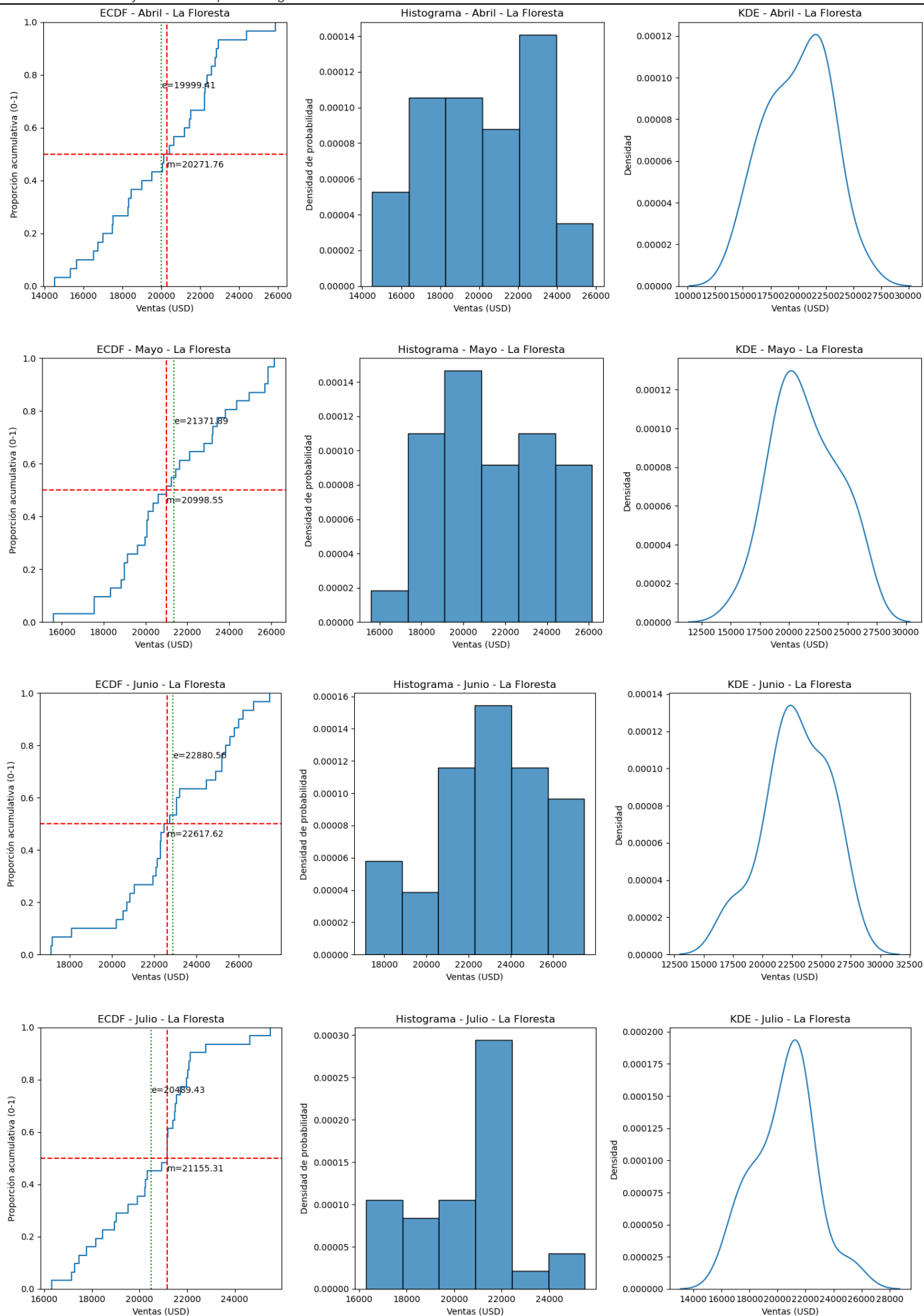
- Carga del archivo excel de datos en data frames de pandas en Python.
- Agrupamiento de los datos de cada una de las tiendas por:
  - 1) Mes del año (*Enero .. Diciembre*)
  - 2) Día de la semana (*Lunes ... Domingo*)

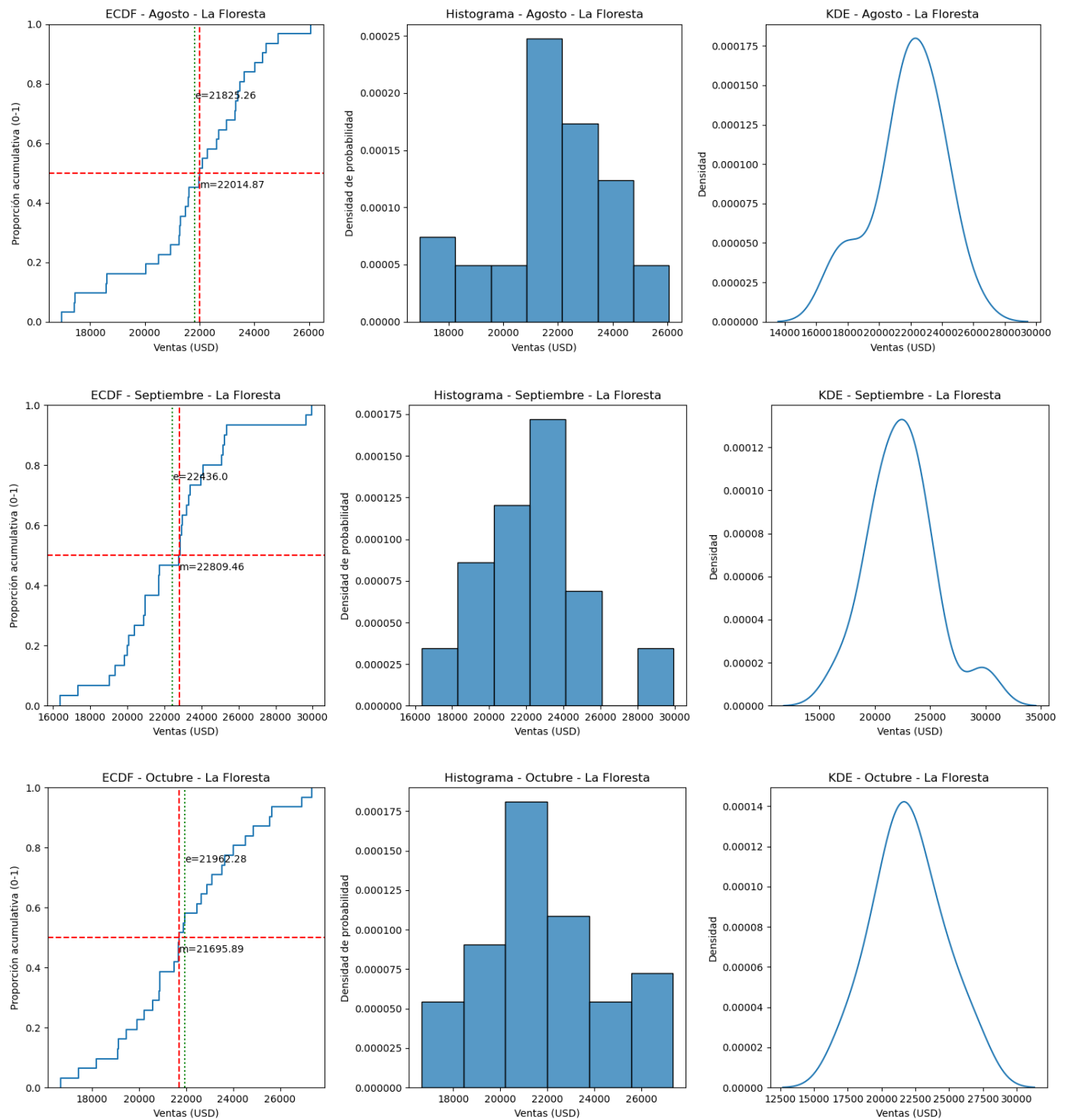
## Análisis de los datos

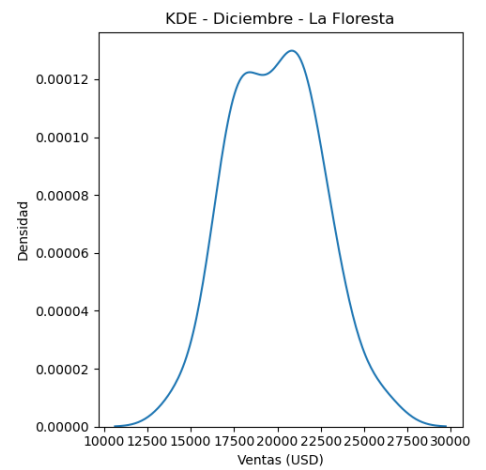
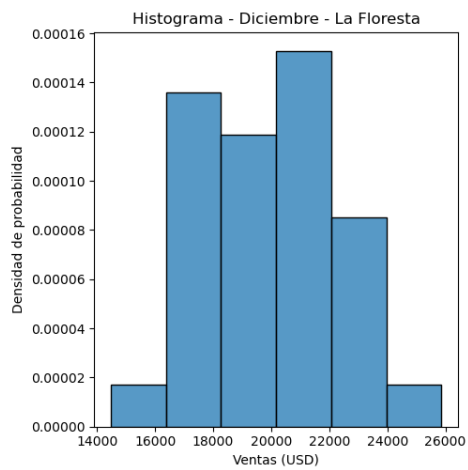
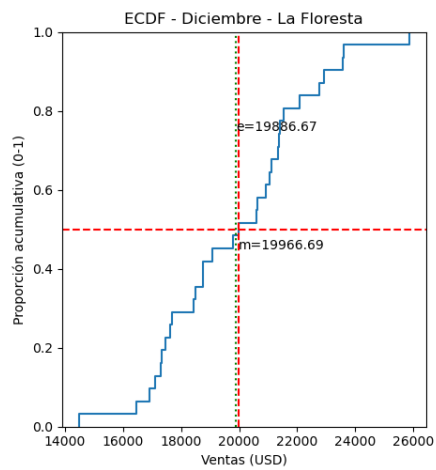
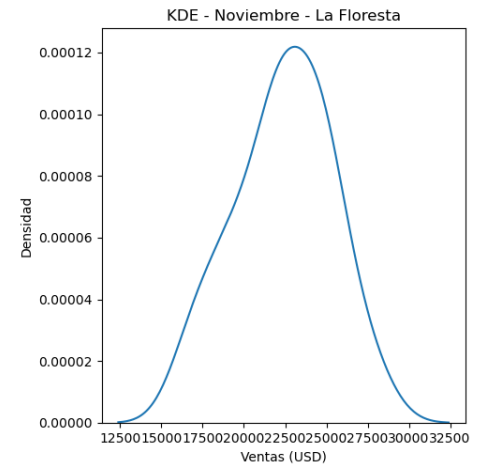
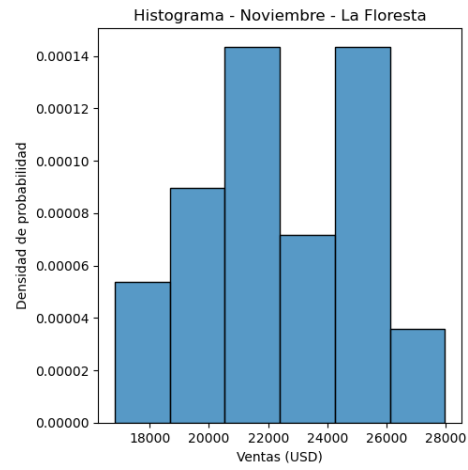
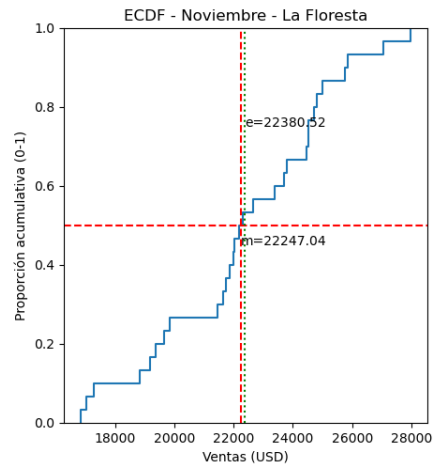
Se determinan funciones empíricas de distribución para las ventas y aproximaciones a las funciones de densidad en cada caso.

### Por mes del año (La Floresta)

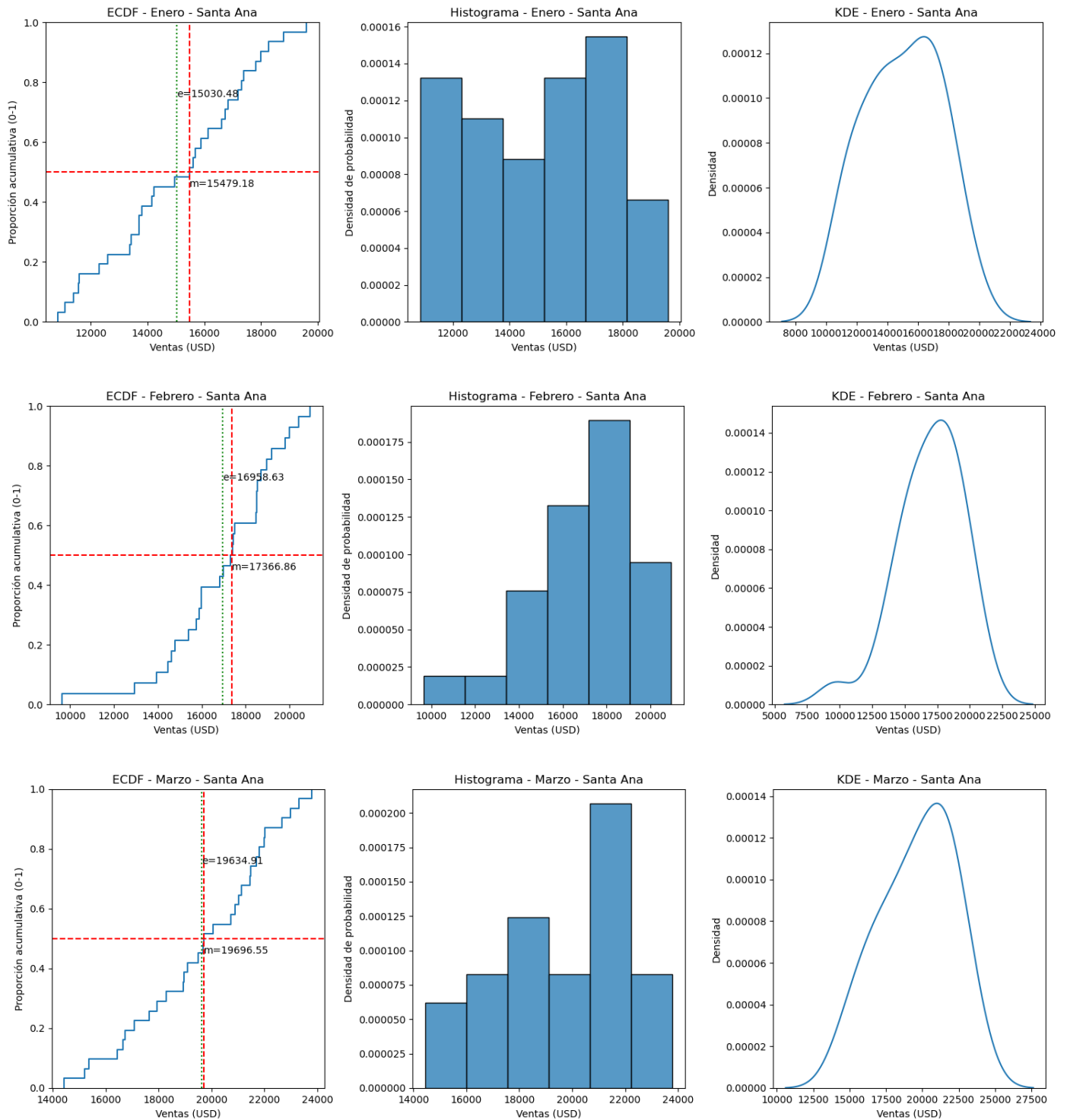




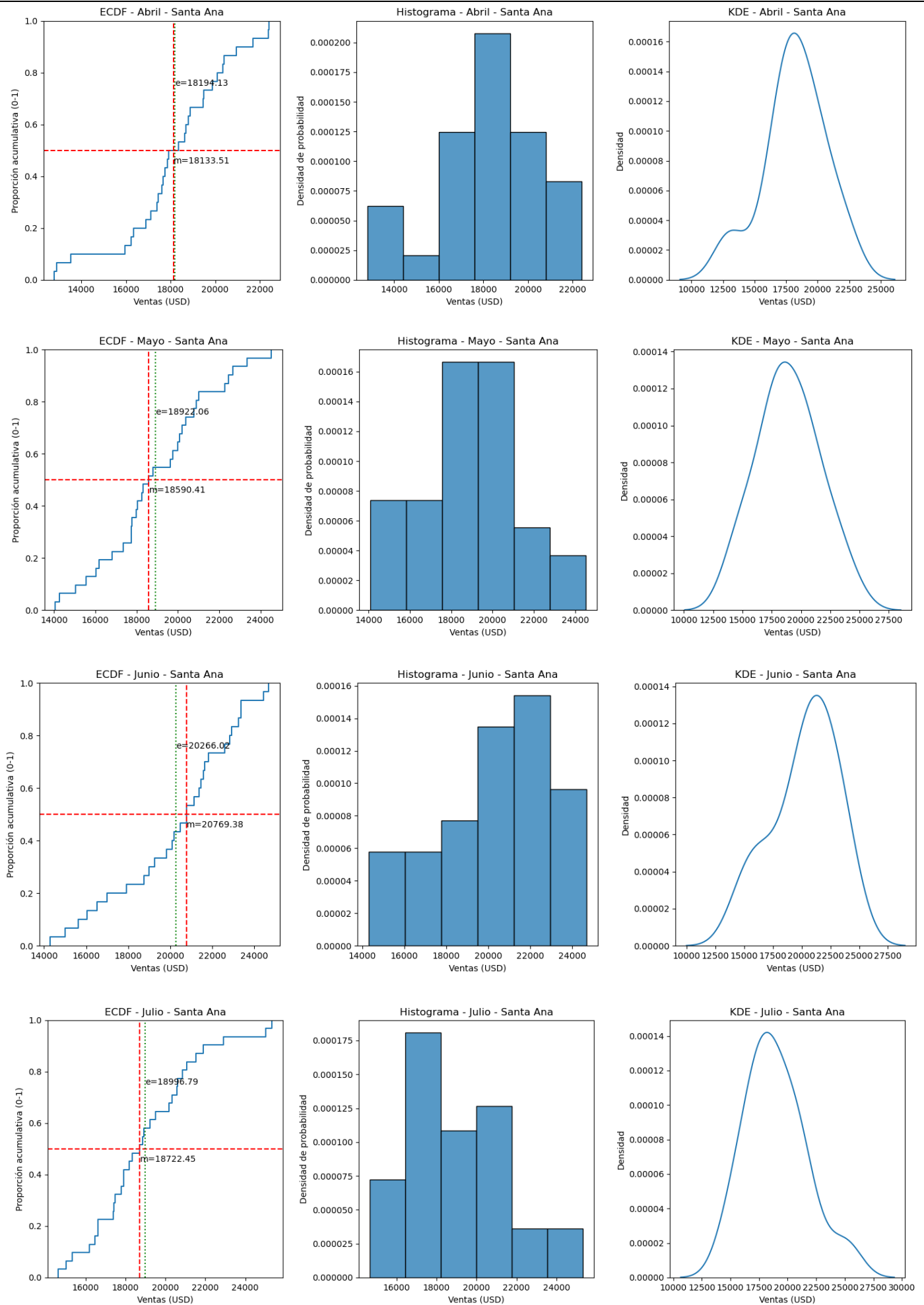


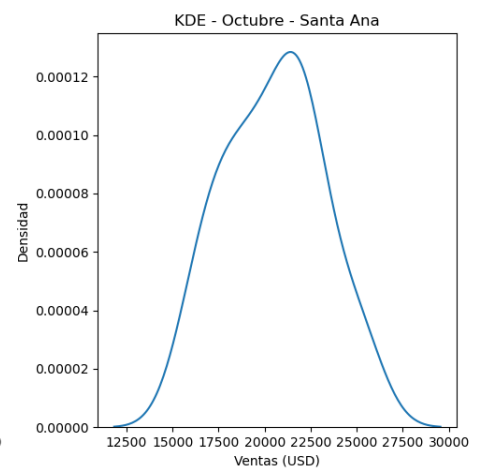
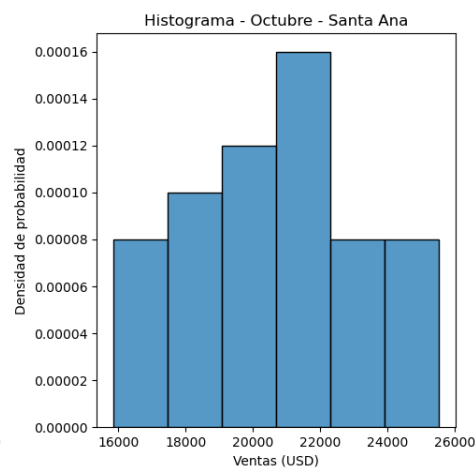
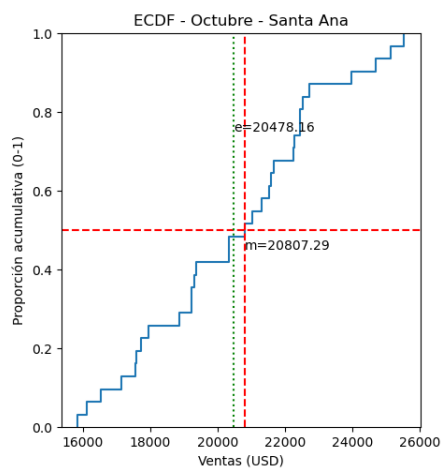
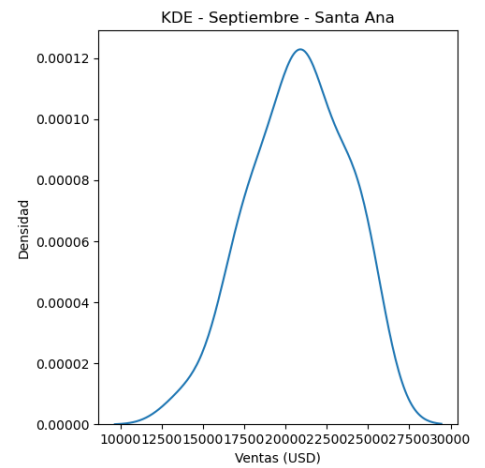
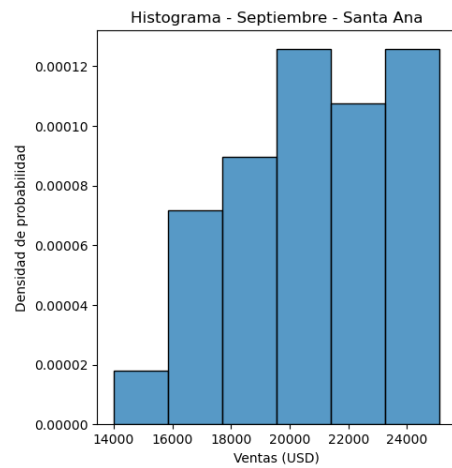
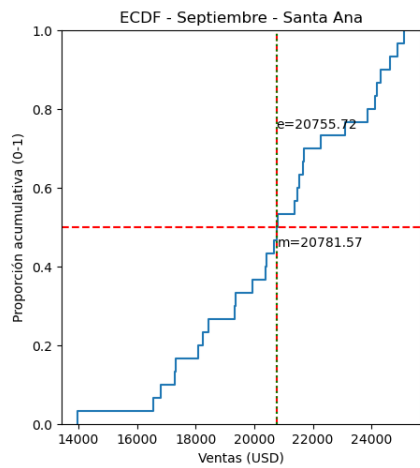
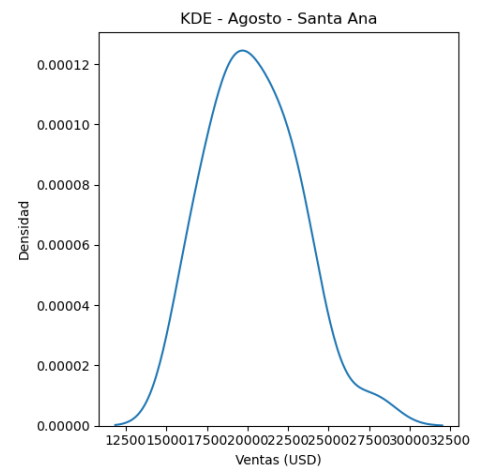
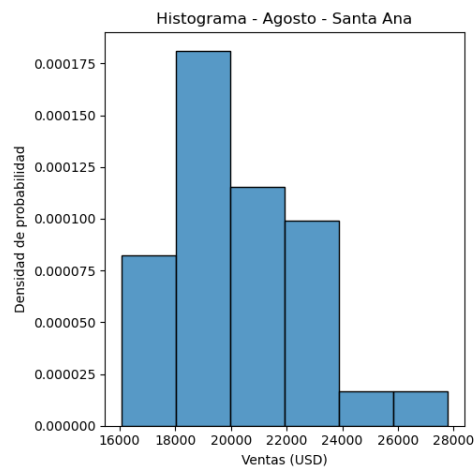
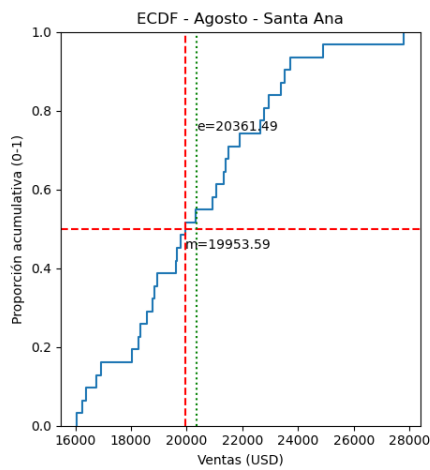


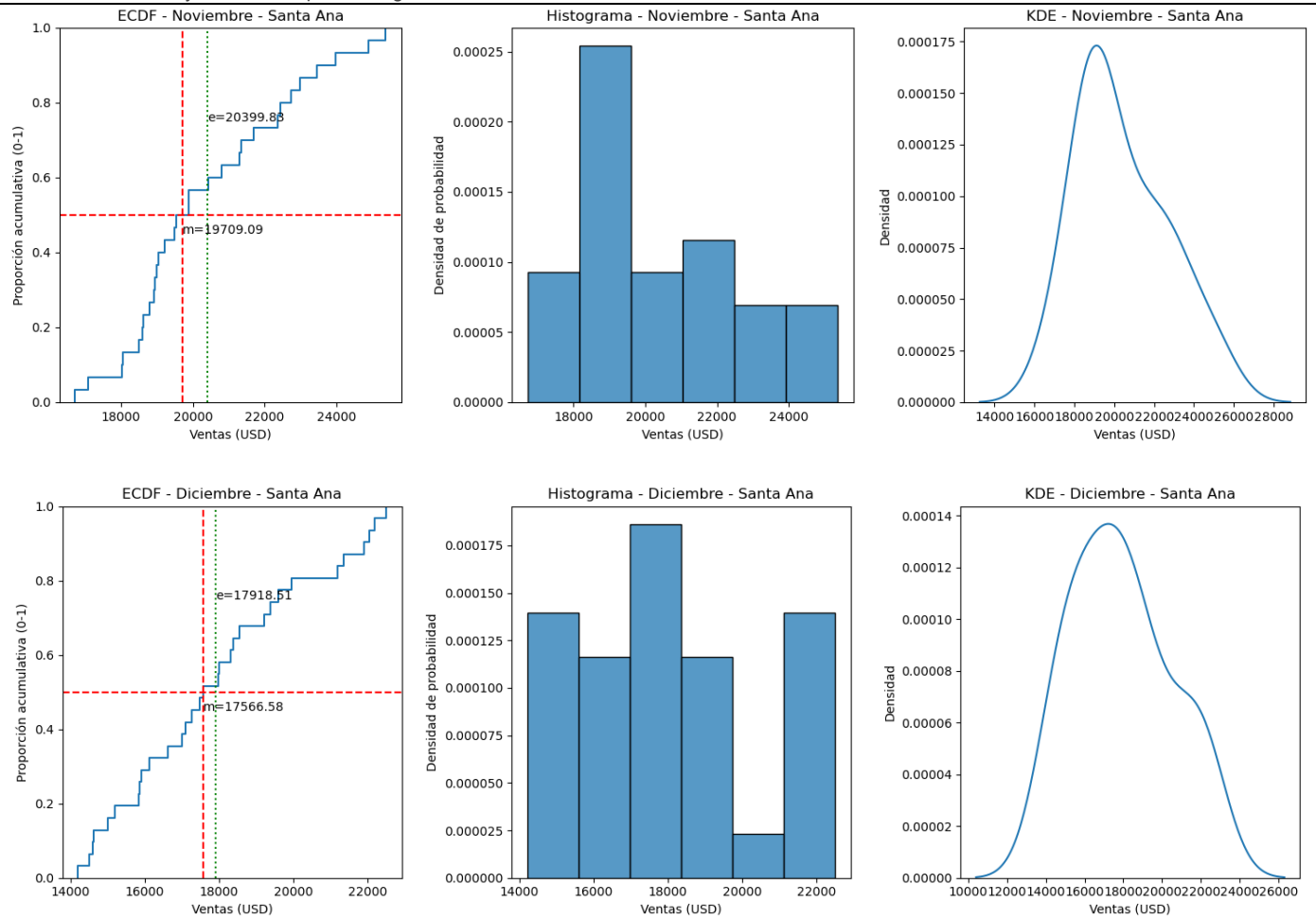
## Por mes del año (Santa Ana)



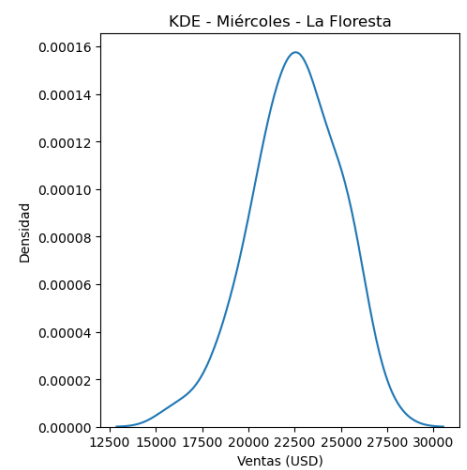
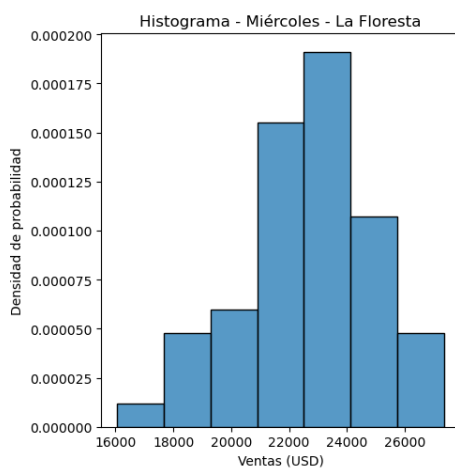
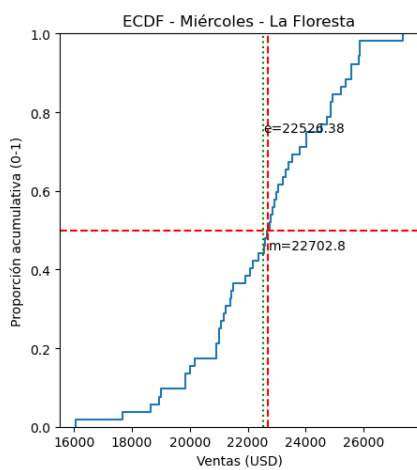
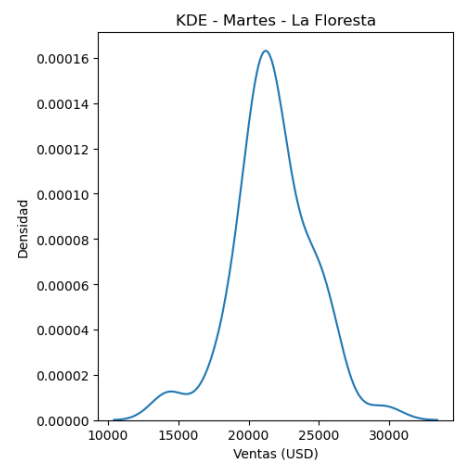
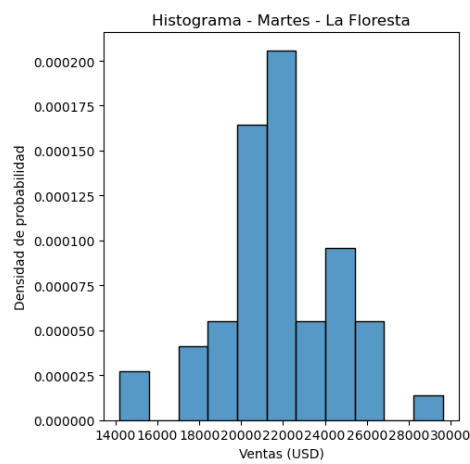
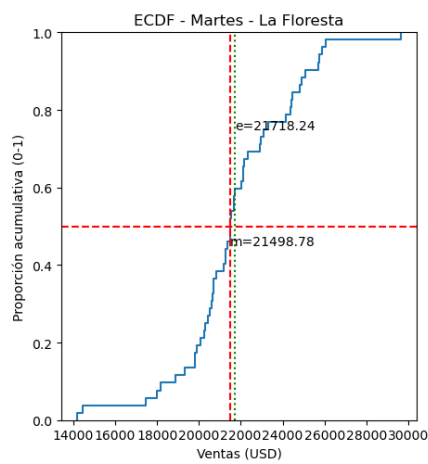
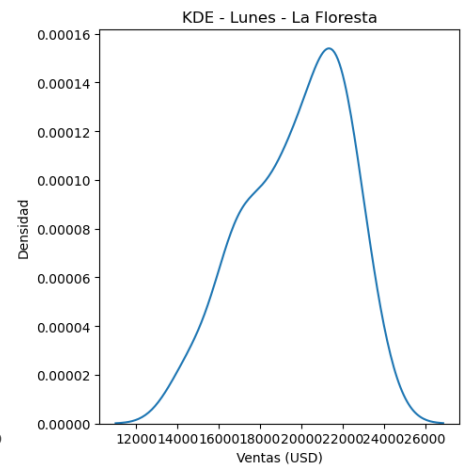
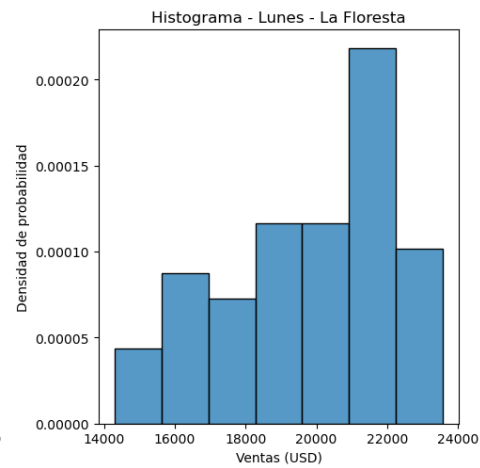
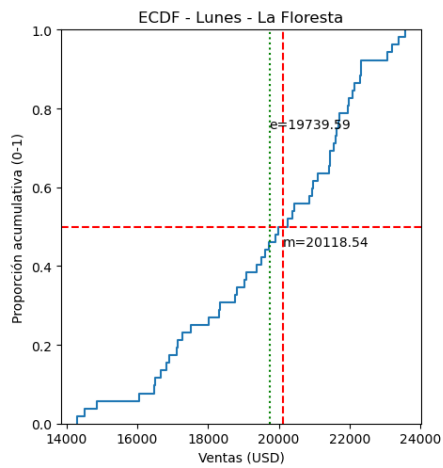


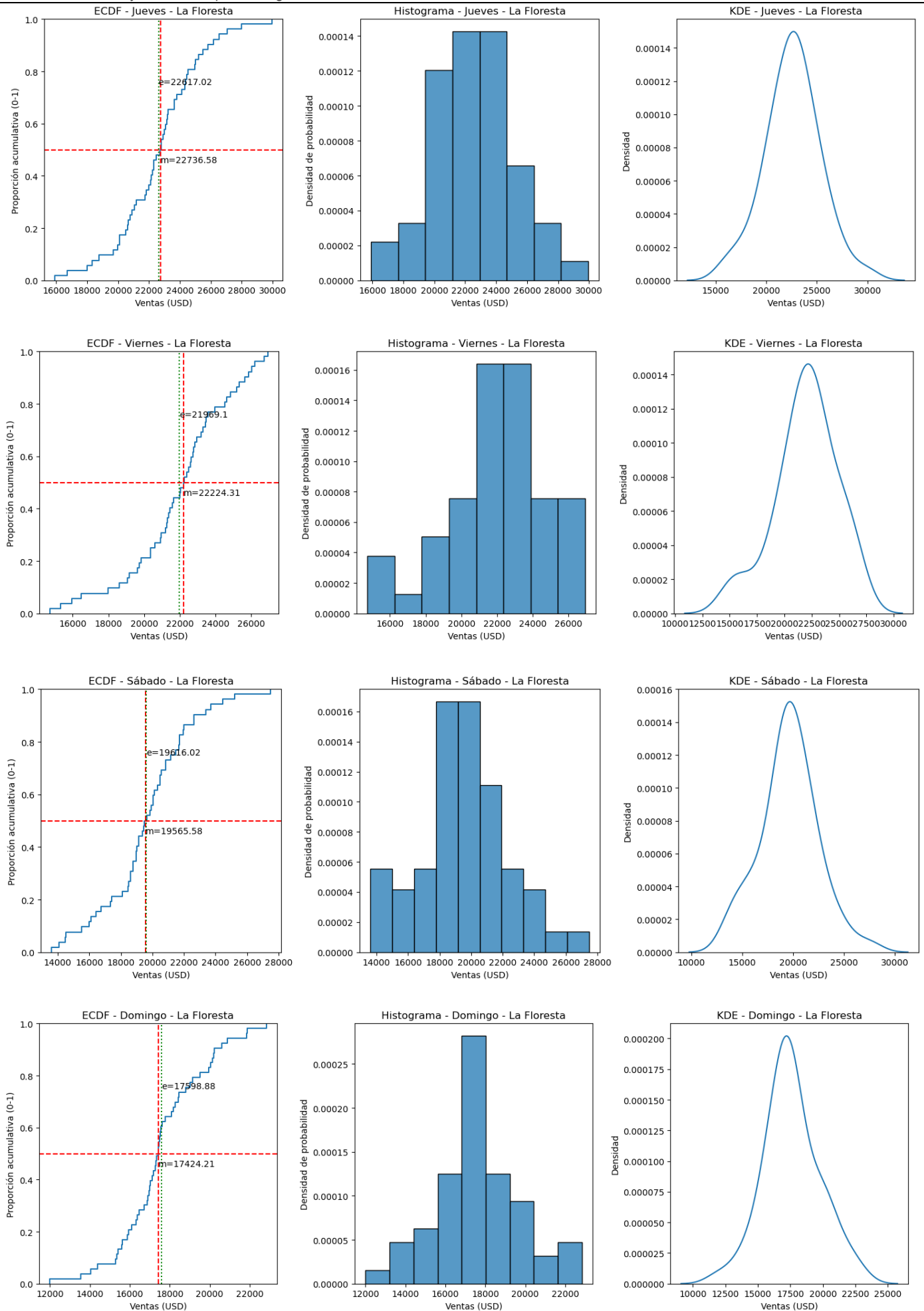




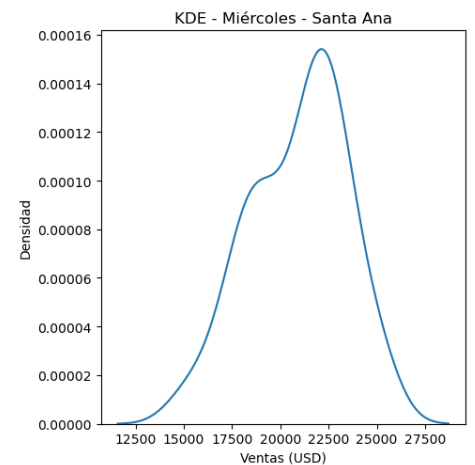
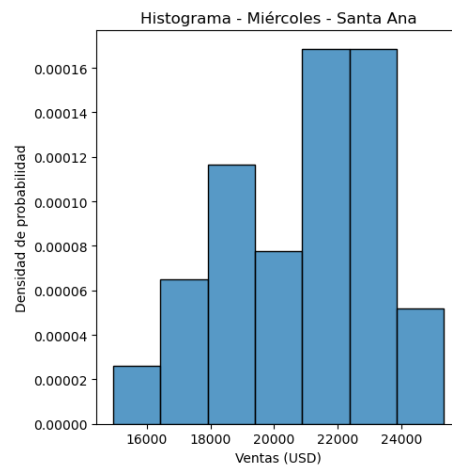
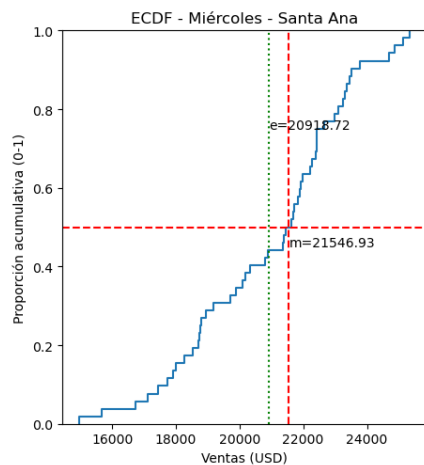
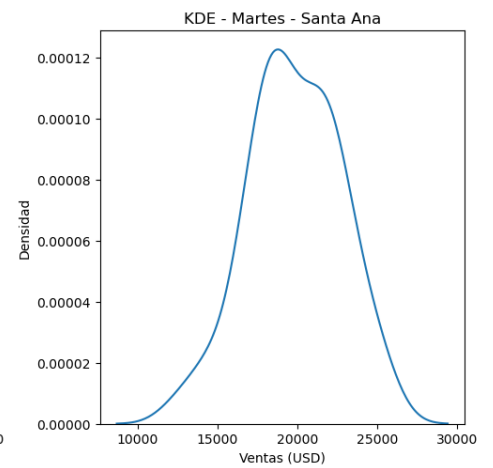
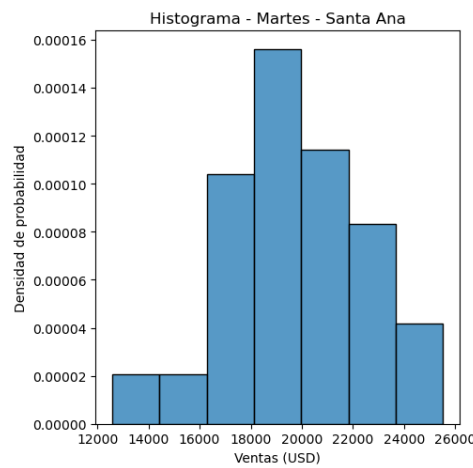
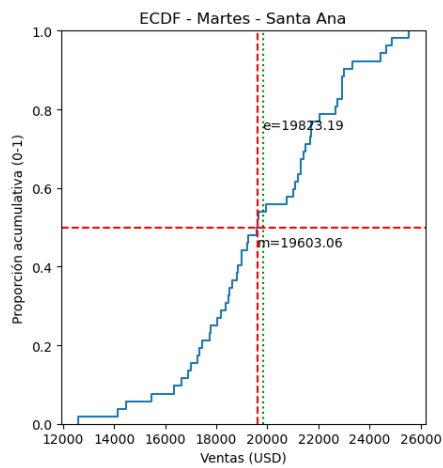
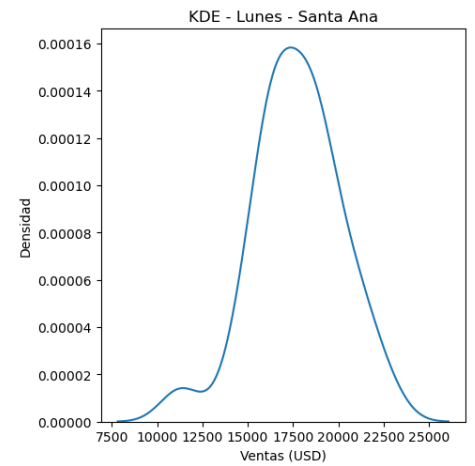
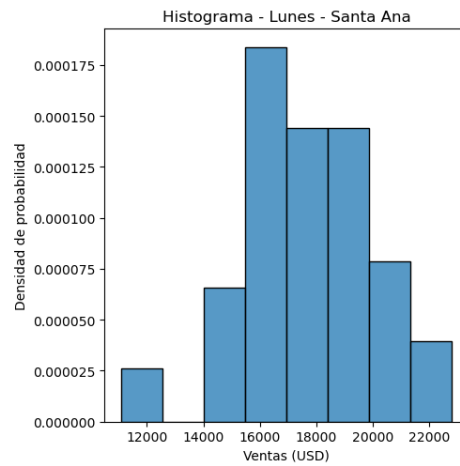
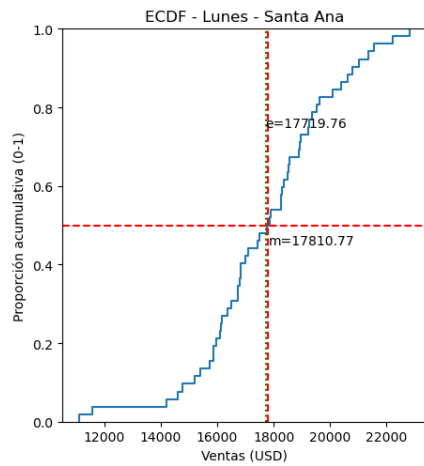


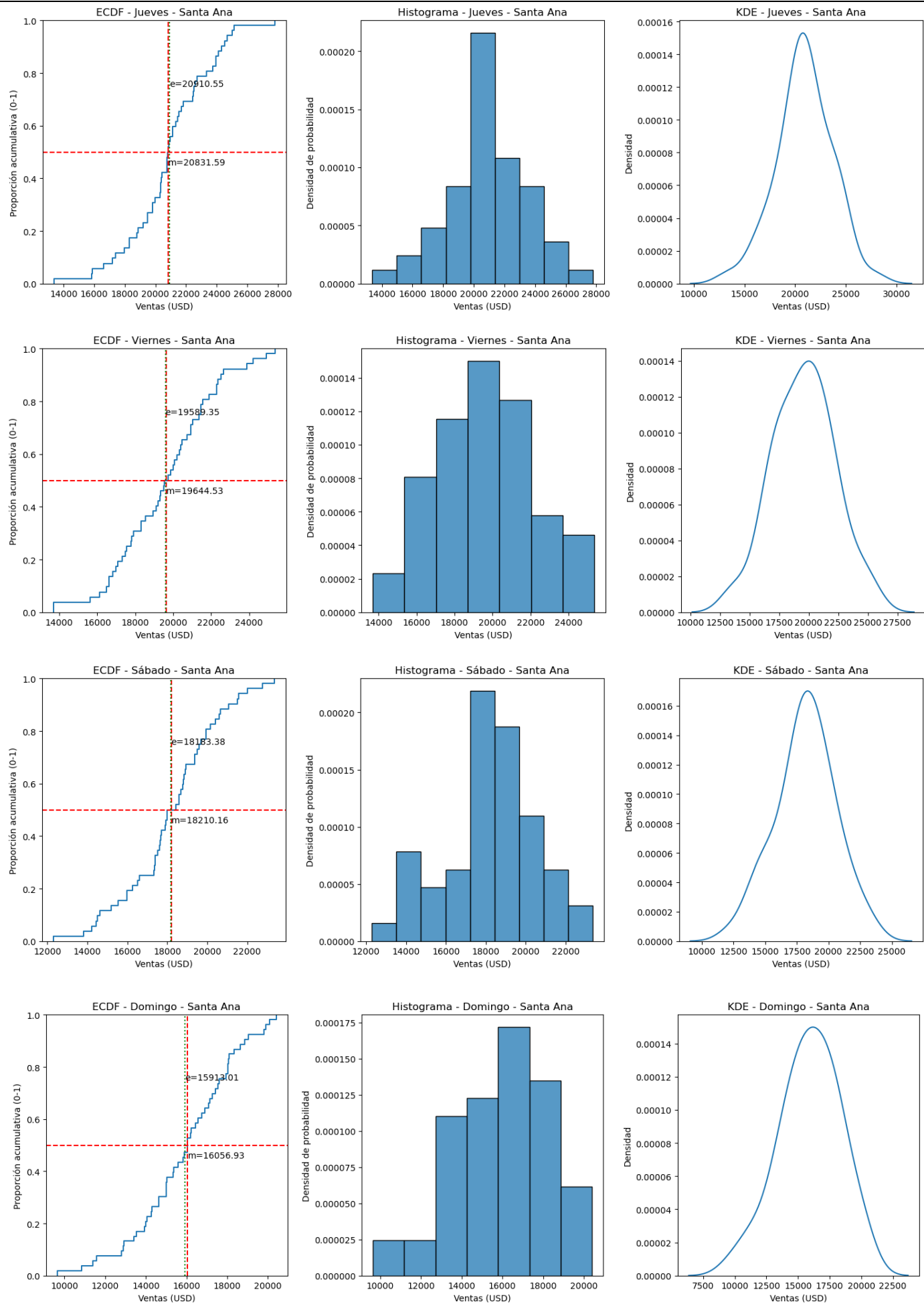
## Por día de la semana (La Floresta)



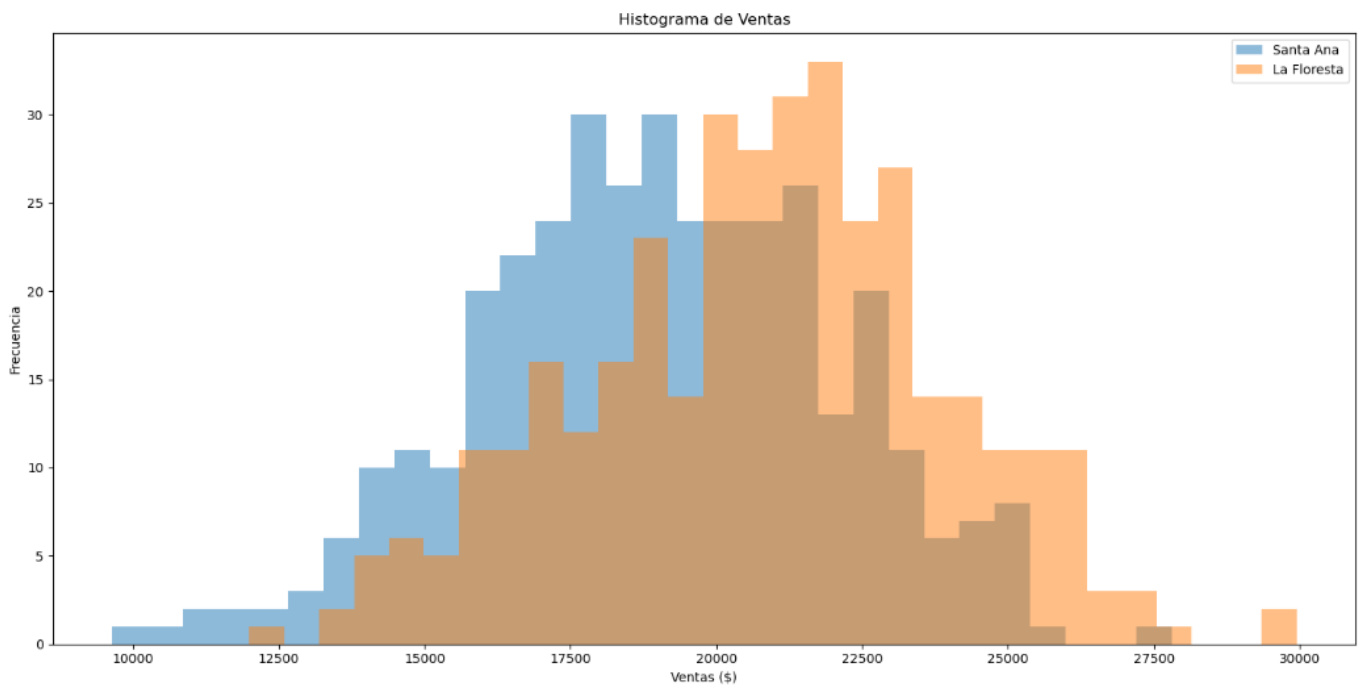
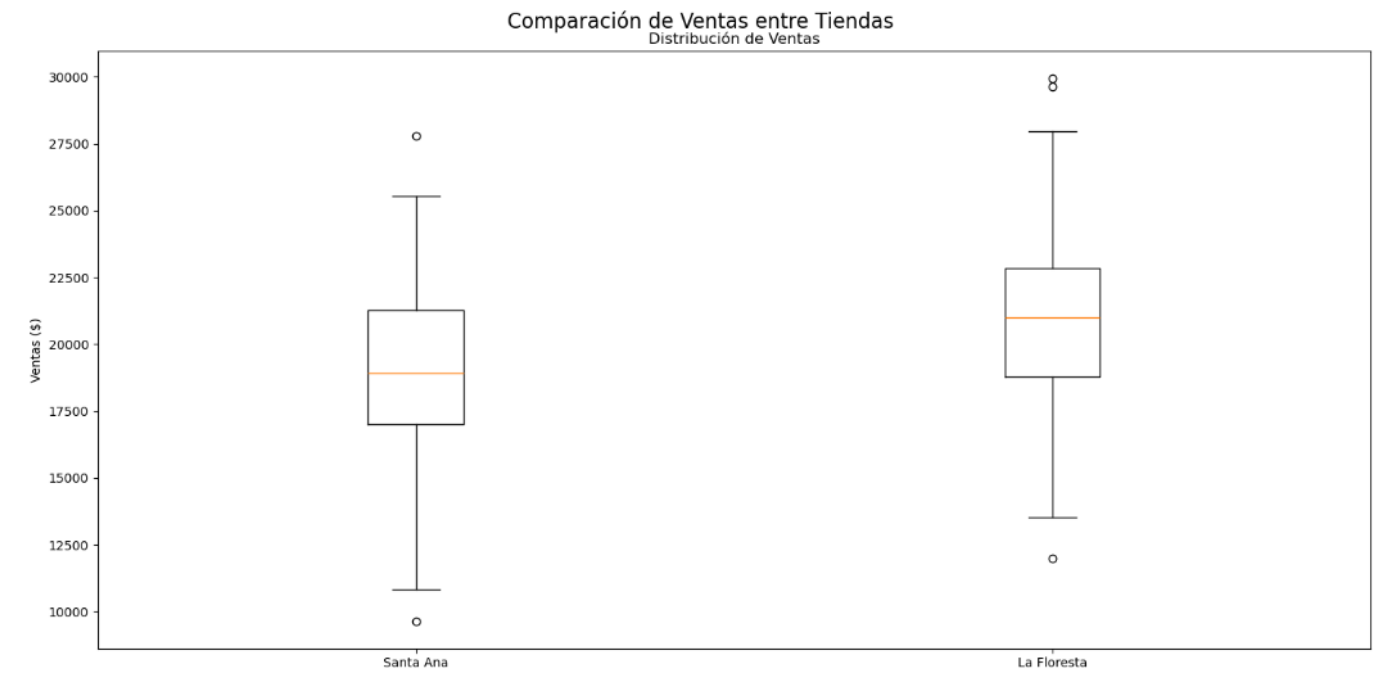


## Por día de la semana (Santa Ana)





## Comparación entre tiendas





## Intervalos de confianza

Se determinan intervalos de confianza para la media de las ventas en cada caso.

Se consideran significancias del 95% y 99%.

### La Floresta – Por mes del año

<b>La Floresta - Ventas enero (USD)</b>				<b>La Floresta - Ventas julio (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	15863.27	16718.61	17870.35	<b>95%</b>	19708.30	20489.43	21541.26
<b>99%</b>	15566.87	16718.61	17870.35	<b>99%</b>	19437.61	20489.43	21541.26
<b>La Floresta - Ventas febrero (USD)</b>				<b>La Floresta - Ventas agosto (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	17768.69	18759.64	20097.76	<b>95%</b>	21002.64	21825.26	22932.95
<b>99%</b>	17421.52	18759.64	20097.76	<b>99%</b>	20717.57	21825.26	22932.95
<b>La Floresta - Ventas marzo (USD)</b>				<b>La Floresta - Ventas septiembre (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	20002.89	21044.76	22447.68	<b>95%</b>	21313.20	22436.0	23949.23
<b>99%</b>	19641.85	21044.76	22447.68	<b>99%</b>	20922.78	22436.0	23949.23
<b>La Floresta - Ventas abril (USD)</b>				<b>La Floresta - Ventas octubre (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	18934.68	19999.41	21434.36	<b>95%</b>	21000.44	21962.28	23257.44
<b>99%</b>	18564.46	19999.41	21434.36	<b>99%</b>	20667.12	21962.28	23257.44
<b>La Floresta - Ventas mayo (USD)</b>				<b>La Floresta - Ventas noviembre (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	20355.83	21371.89	22740.03	<b>95%</b>	21290.80	22380.52	23849.14
<b>99%</b>	20003.74	21371.89	22740.03	<b>99%</b>	20911.89	22380.52	23849.14
<b>La Floresta - Ventas junio (USD)</b>				<b>La Floresta - Ventas diciembre (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	21867.75	22880.56	24245.53	<b>95%</b>	18954.06	19886.67	21142.46
<b>99%</b>	21515.59	22880.56	24245.53	<b>99%</b>	18630.88	19886.67	21142.46

## Santa Ana – Por mes del año

### Santa Ana - Ventas enero (USD)

	Límite inferior	Media	Límite superior
95%	14120.01	15030.48	16256.45
99%	13804.50	15030.48	16256.45

### Santa Ana - Ventas febrero (USD)

	Límite inferior	Media	Límite superior
95%	15986.18	16958.63	18271.77
99%	15645.48	16958.63	18271.77

### Santa Ana - Ventas marzo (USD)

	Límite inferior	Media	Límite superior
95%	18702.64	19634.91	20890.26
99%	18379.57	19634.91	20890.26

### Santa Ana - Ventas abril (USD)

	Límite inferior	Media	Límite superior
95%	17284.58	18194.13	19419.96
99%	16968.31	18194.13	19419.96

### Santa Ana - Ventas mayo (USD)

	Límite inferior	Media	Límite superior
95%	17953.51	18922.06	20226.24
99%	17617.87	18922.06	20226.24

### Santa Ana - Ventas junio (USD)

	Límite inferior	Media	Límite superior
95%	19210.47	20266.02	21688.59
99%	18843.45	20266.02	21688.59

### Santa Ana - Ventas julio (USD)

	Límite inferior	Media	Límite superior
95%	18028.10	18996.79	20301.18
99%	17692.41	18996.79	20301.18

### Santa Ana - Ventas agosto (USD)

	Límite inferior	Media	Límite superior
95%	19340.83	20361.49	21735.85
99%	18987.13	20361.49	21735.85

### Santa Ana - Ventas septiembre (USD)

	Límite inferior	Media	Límite superior
95%	19694.35	20755.72	22186.13
99%	19325.30	20755.72	22186.13

### Santa Ana - Ventas octubre (USD)

	Límite inferior	Media	Límite superior
95%	19502.91	20478.16	21791.38
99%	19164.94	20478.16	21791.38

### Santa Ana - Ventas noviembre (USD)

	Límite inferior	Media	Límite superior
95%	19548.91	20399.83	21546.62
99%	19253.04	20399.83	21546.62

### Santa Ana - Ventas diciembre (USD)

	Límite inferior	Media	Límite superior
95%	16996.23	17918.51	19160.39
99%	16676.62	17918.51	19160.39

## La Floresta – Por día de la semana

## La Floresta - Ventas LUNES (USD)

	Límite inferior	Media	Límite superior
<b>95%</b>	19061.86	19739.59	20642.87
<b>99%</b>	18836.31	19739.59	20642.87

## La Floresta - Ventas MARTES (USD)

	Límite inferior	Media	Límite superior
<b>95%</b>	20945.26	21718.24	22748.47
<b>99%</b>	20688.01	21718.24	22748.47

## La Floresta - Ventas MIÉRCOLES (USD)

	Límite inferior	Media	Límite superior
<b>95%</b>	21876.93	22526.38	23391.96
<b>99%</b>	21660.79	22526.38	23391.96

## La Floresta - Ventas JUEVES (USD)

	Límite inferior	Media	Límite superior
<b>95%</b>	21862.10	22617.02	23623.19
<b>99%</b>	21610.86	22617.02	23623.19

## La Floresta - Ventas VIERNES (USD)

	Límite inferior	Media	Límite superior
<b>95%</b>	21184.22	21969.1	23015.19
<b>99%</b>	20923.01	21969.1	23015.19

## La Floresta - Ventas SÁBADO (USD)

	Límite inferior	Media	Límite superior
<b>95%</b>	18833.91	19616.02	20658.41
<b>99%</b>	18573.62	19616.02	20658.41

## La Floresta - Ventas DOMINGO (USD)

	Límite inferior	Media	Límite superior
<b>95%</b>	17009.25	17598.88	18384.54
<b>99%</b>	16813.23	17598.88	18384.54

## Santa Ana – Por día de la semana

<b>Santa Ana - Ventas LUNES (USD)</b>				<b>Santa Ana - Ventas VIERNES (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	17052.91	17719.76	18608.54	<b>95%</b>	18866.49	19589.35	20552.78
<b>99%</b>	16830.98	17719.76	18608.54	<b>99%</b>	18625.92	19589.35	20552.78
<b>Santa Ana - Ventas MARTES (USD)</b>				<b>Santa Ana - Ventas SÁBADO (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	19028.07	19823.19	20882.94	<b>95%</b>	17529.91	18183.38	19054.33
<b>99%</b>	18763.44	19823.19	20882.94	<b>99%</b>	17312.43	18183.38	19054.33
<b>Santa Ana - Ventas MIÉRCOLES (USD)</b>				<b>Santa Ana - Ventas DOMINGO (USD)</b>			
	Límite inferior	Media	Límite superior		Límite inferior	Media	Límite superior
<b>95%</b>	20227.91	20918.72	21839.43	<b>95%</b>	15245.30	15913.01	16802.7
<b>99%</b>	19998.01	20918.72	21839.43	<b>99%</b>	15023.32	15913.01	16802.7
<b>Santa Ana - Ventas JUEVES (USD)</b>							
	Límite inferior	Media	Límite superior				
<b>95%</b>	20155.03	20910.55	21917.51				
<b>99%</b>	19903.59	20910.55	21917.51				

## Pruebas de Hipótesis

La intuición de Don Francisco es que en la tienda “La Floresta” se vende menos que en “Santa Ana”.

Por lo tanto, utilizamos la media de las ventas basad

$H_0$  (Hipótesis nula)

No hay diferencia significativa en las ventas de ambas tiendas.

Por lo tanto, la media de las ventas diarias de ambas tiendas no es significativamente diferente.

$H_1$  (Hipótesis alternativa)

En 'La Floresta' se vende menos que en 'Santa Ana'.

Por lo tanto, la media de las ventas diarias de 'La Floresta' es menor que la media de las ventas diarias de 'Santa Ana'.

Utilizamos niveles de significancia del 95% y del 99%.

En ambos casos, concluimos que no hay evidencia significativa para rechazar la hipótesis nula.

## Conclusiones

### a) Comportamiento de ventas por mes:

- Para ambas tiendas se observan menores ventas para el mes de **enero**.
- Para la tienda Santa Ana se observa que el mes de mayor venta es **septiembre**.
- Para la tienda La Floresta se observa que el mes de mayor venta es **noviembre**.

### b) Comportamiento de ventas por día de la semana:

- Para ambas tiendas se observa que el día con menores ventas es el día **domingo**
- Para ambas tiendas se observa que el día con mayores ventas es el día **jueves**.

### c) Comparación entre tiendas:

*(Aquí se interpreta el resultado de la prueba t, indicando si hay una diferencia significativa en las ventas entre ambas tiendas)*

- **Para significancias del 95%:** No hay evidencia suficiente para concluir que la tienda 'La Floresta' vende menos que la tienda 'Santa Ana'.
- **Para significancia del 99%:** No hay evidencia suficiente para concluir que la tienda 'La Floresta' vende menos que la tienda 'Santa Ana'.

### d) Recomendaciones para Don Francisco:

- Mes recomendado para vacaciones:  
*Enero\*\* \*(basado en el mes con menores ventas en ambas tiendas)*
- Mes recomendado para inversiones:  
*Septiembre y Noviembre\*\* \*(basado en los meses con mayores ventas para Santa Ana y La Floresta respectivamente)*
- Ajustes en los horarios de atención y contratación de personal:  
Tanto en La Floresta como en Santa Ana deberían reducir el personal los días **domingo** que es el día de la semana que de media tienen menores ventas ambas tiendas.  
Tanto en La Floresta como en Santa Ana debería reforzar el personal los días **jueves** que es el día de la semana que de media tienen mayores ventas ambas tiendas *(basado en el análisis por día de la semana)*

## Anexo

Archivo *Examen\_final\_PEIA\_Grupo5.ipynb* con el código python que respalda los cálculos y conclusiones.