



Facultad de Ingeniería

Carrera de Especialización en Inteligencia Artificial

Examen Final

Probabilidad y Estadística para la Inteligencia Artificial

Autores (Grupo 9):
Hagata, Juan Pablo
Juarez, Yelamos, Fausto
Belen, Santiago Francisco

Docentes:
Argoty, Camilo
Dreszman, Alan

Enunciado

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%

Planteamiento de los problemas

A) **Problema de Investigación 1:** Análisis de ventas mensuales de las tiendas "Santa Ana" y "La Floresta".

1. Calcular y graficar las funciones de distribución empírica para cada mes de ambas tiendas.
2. Estimar y graficar las funciones de densidad utilizando el método de kernels.
3. Construir los intervalos de confianza empíricos al 95% y 99% para las ventas mensuales.
4. Evaluar la implicancia de los resultados en la selección de tiempos de vacaciones y momentos óptimos para inversiones.

B) **Problema de Investigación 2:** Análisis de ventas diarias de las tiendas "Santa Ana" y "La Floresta".

1. Calcular y graficar las funciones de distribución empírica para cada día de la semana de ambas tiendas.
2. Estimar y graficar las funciones de densidad utilizando el método de kernels.
3. Construir los intervalos de confianza empíricos al 95% y 99% para las ventas diarias.
4. Evaluar la implicancia de los resultados en la gestión de recursos humanos.

C) **Problema de Investigación 3:** Análisis de comparación de ventas entre las tiendas "Santa Ana" y "La Floresta".

1. Realizar una prueba de hipótesis para comparar las ventas promedio entre 'Santa Ana' y 'La Floresta'.
2. Calcular los valores p para los niveles de significancia del 95.
3. Evaluar si las percepciones de Don Francisco son estadísticamente válidas.

Procesamiento de datos realizados

Con el objetivo de convertir y organizar la información de las fechas para facilitar análisis posteriores sobre cómo las ventas varían mensual y semanalmente se realizaron los siguientes pasos de procesamiento:

A) Conversión de Fechas a Formato DateTime:

Se utilizó la función `pd.to_datetime()` de la biblioteca **pandas** para convertir la columna **Fecha** en ambos conjuntos de datos. Este paso es crucial porque transforma las fechas, originalmente en un formato no estandarizado, a un formato de fecha y hora reconocible por Python, lo que permite realizar operaciones de fechas de manera más eficiente, como extraer el mes o el día de la semana directamente.

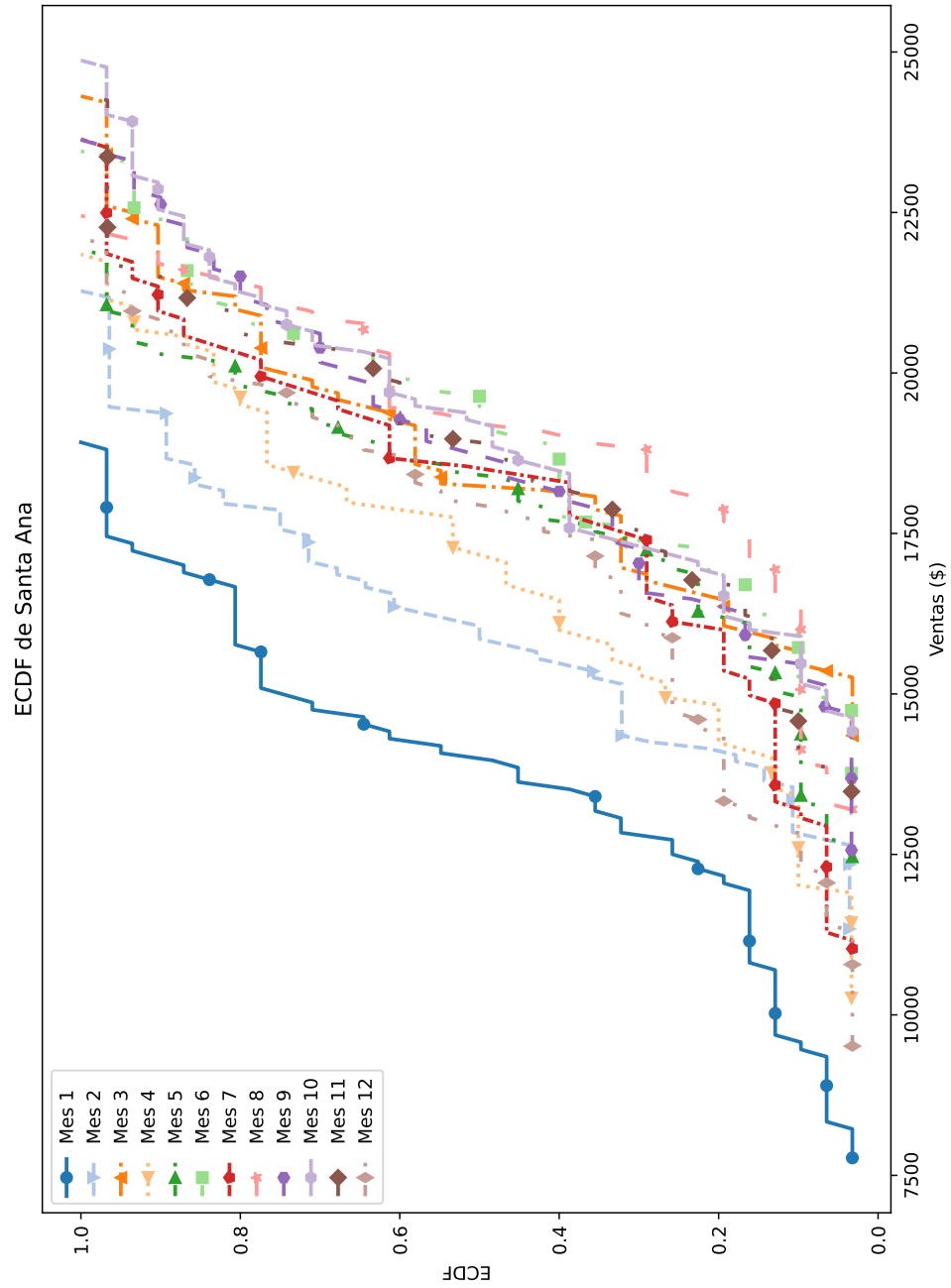
B) Extracción del Mes y Día de la Semana:

Se utilizaron los atributos `.dt.month` y `.dt.day_name()` de la biblioteca **pandas** para extraer los datos del mes y el día de la columna Fecha. Los mismos, fueron agregados nuevas columnas del DataFrame: Mes y DiaSemana, respectivamente.

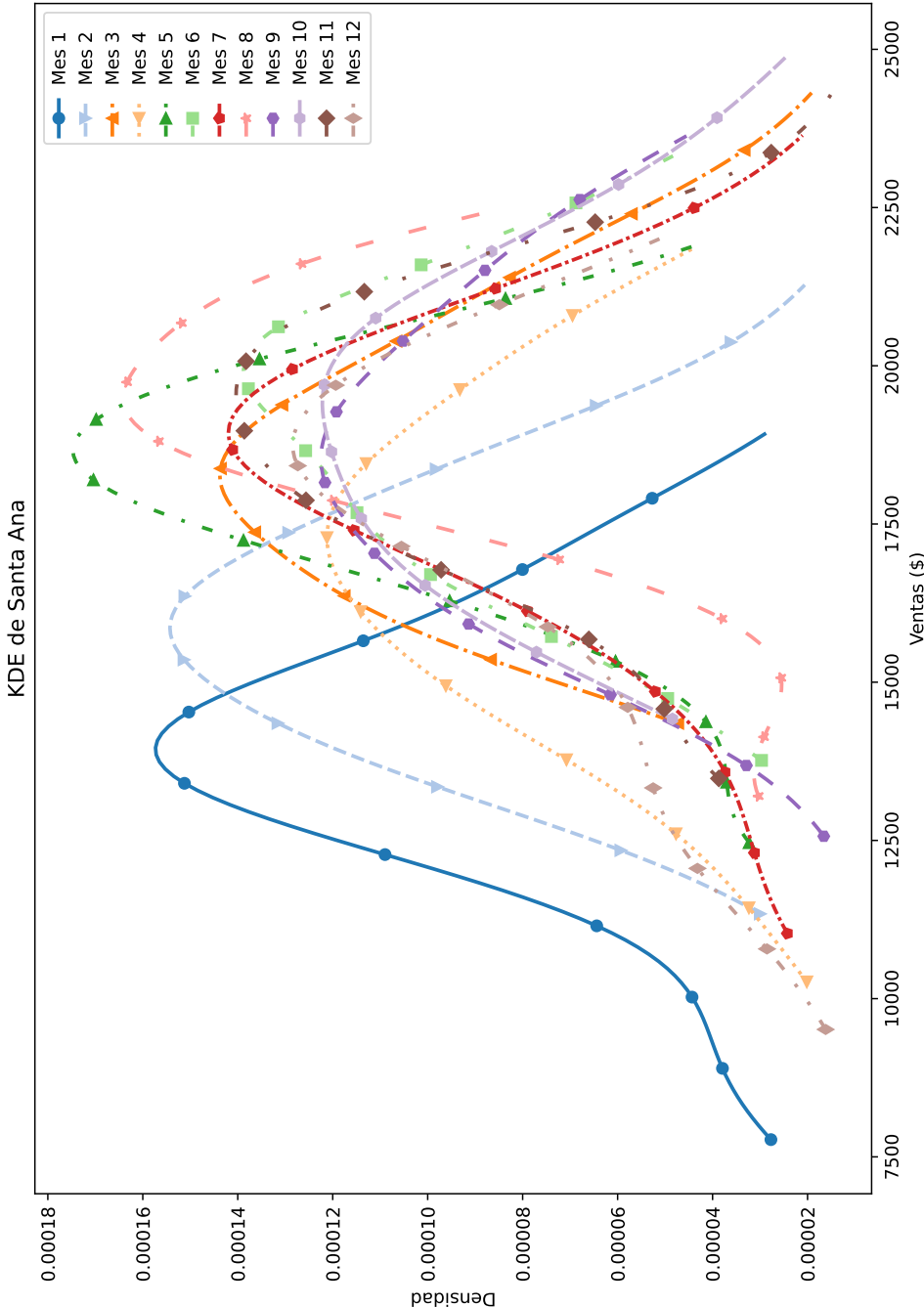
```
1 # Convertir la columna de fechas a datetime
2 santa_ana['Fecha'] = pd.to_datetime(santa_ana['Fecha'])
3 la_floresta['Fecha'] = pd.to_datetime(la_floresta['Fecha'])
4
5 # Agregar columnas para mes y día de la semana
6 santa_ana['Mes'] = santa_ana['Fecha'].dt.month
7 santa_ana['DiaSemana'] = santa_ana['Fecha'].dt.day_name()
8
9 la_floresta['Mes'] = la_floresta['Fecha'].dt.month
10 la_floresta['DiaSemana'] = la_floresta['Fecha'].dt.day_name()
11
```

Resultados

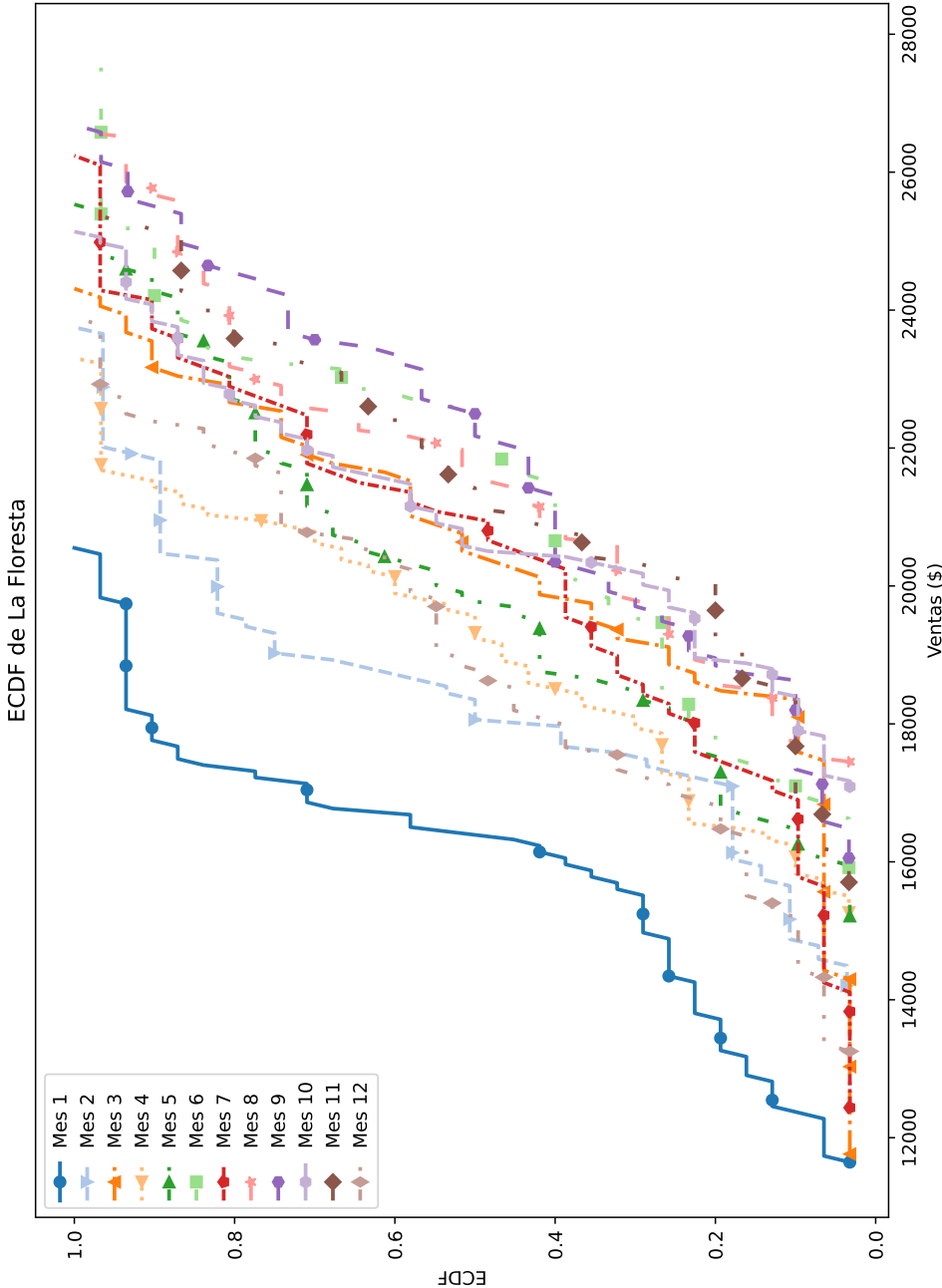
ECDF Santa Ana - MENSUAL -



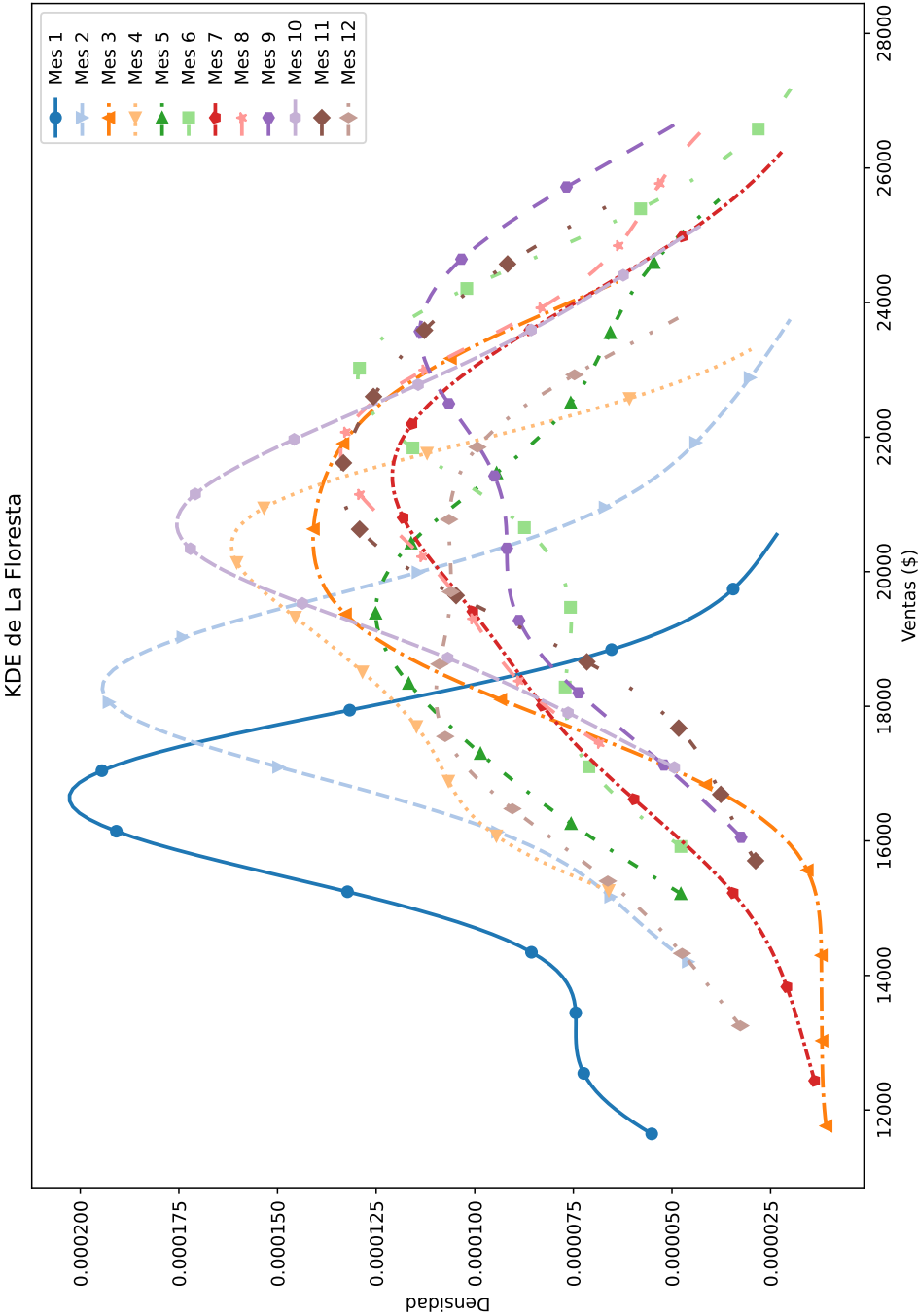
KDE Santa Ana - MENSUAL -



ECDF La Floresta - MENSUAL -

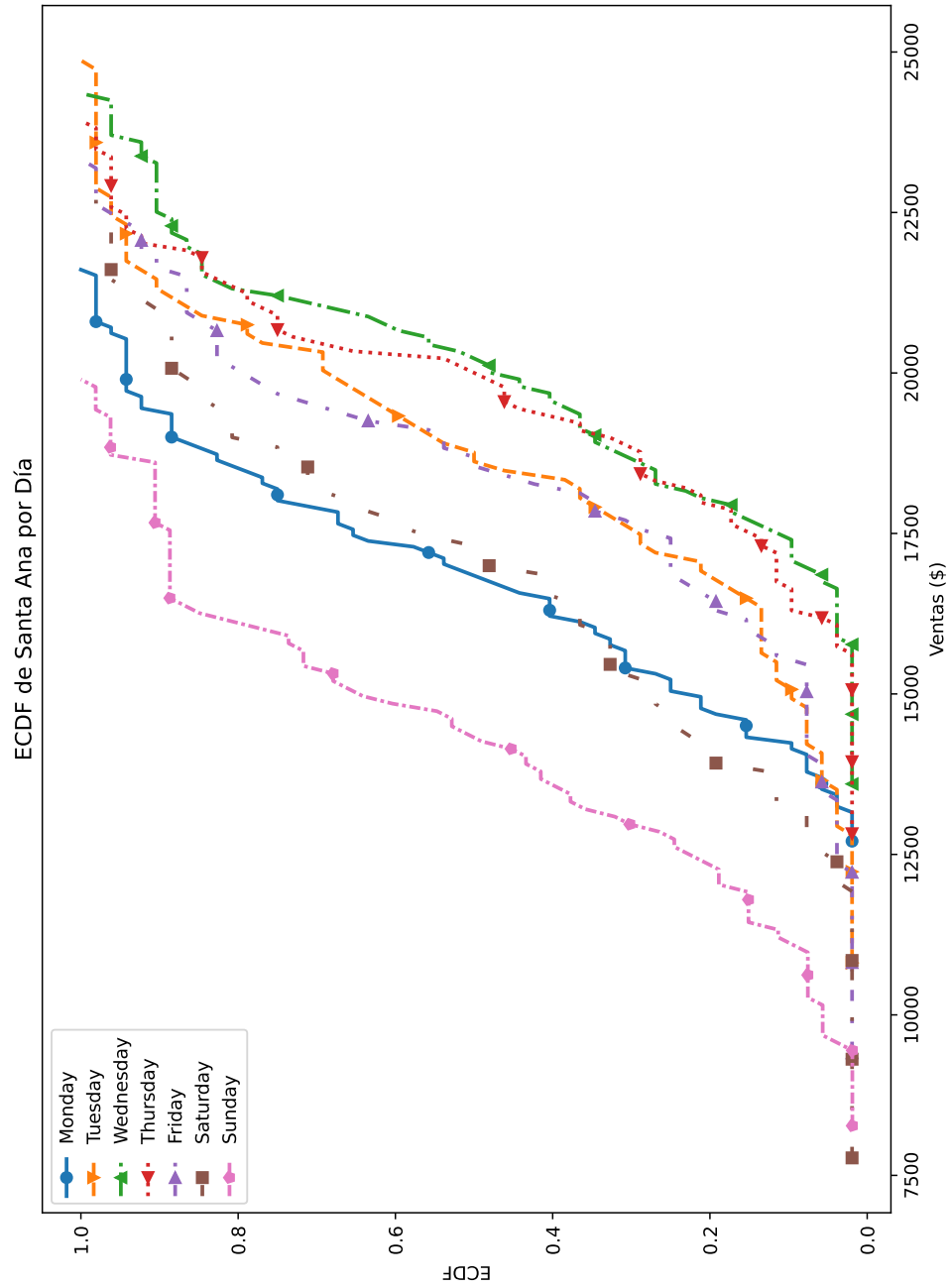


KDE La Floresta - MENSUAL -

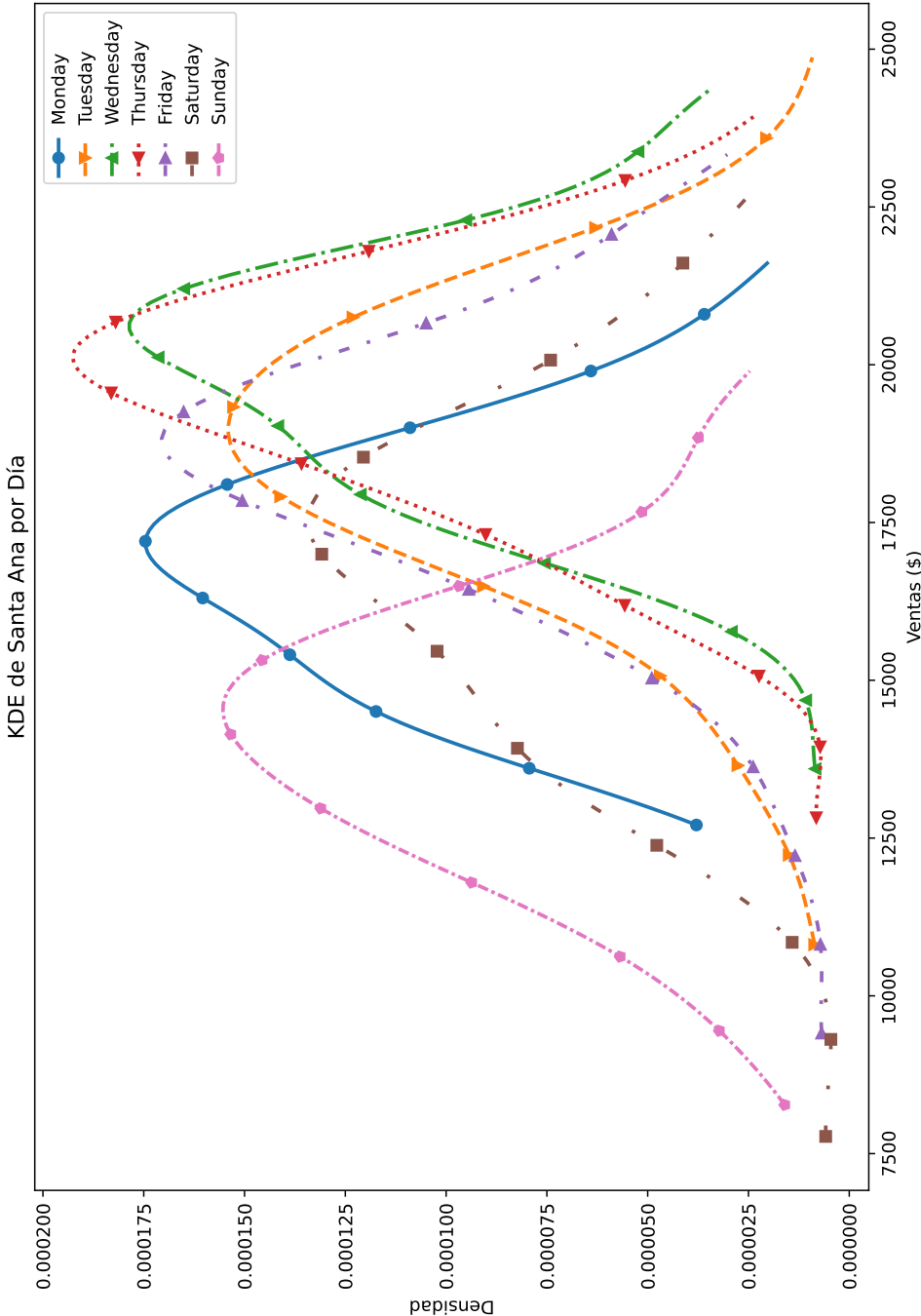


Resultados

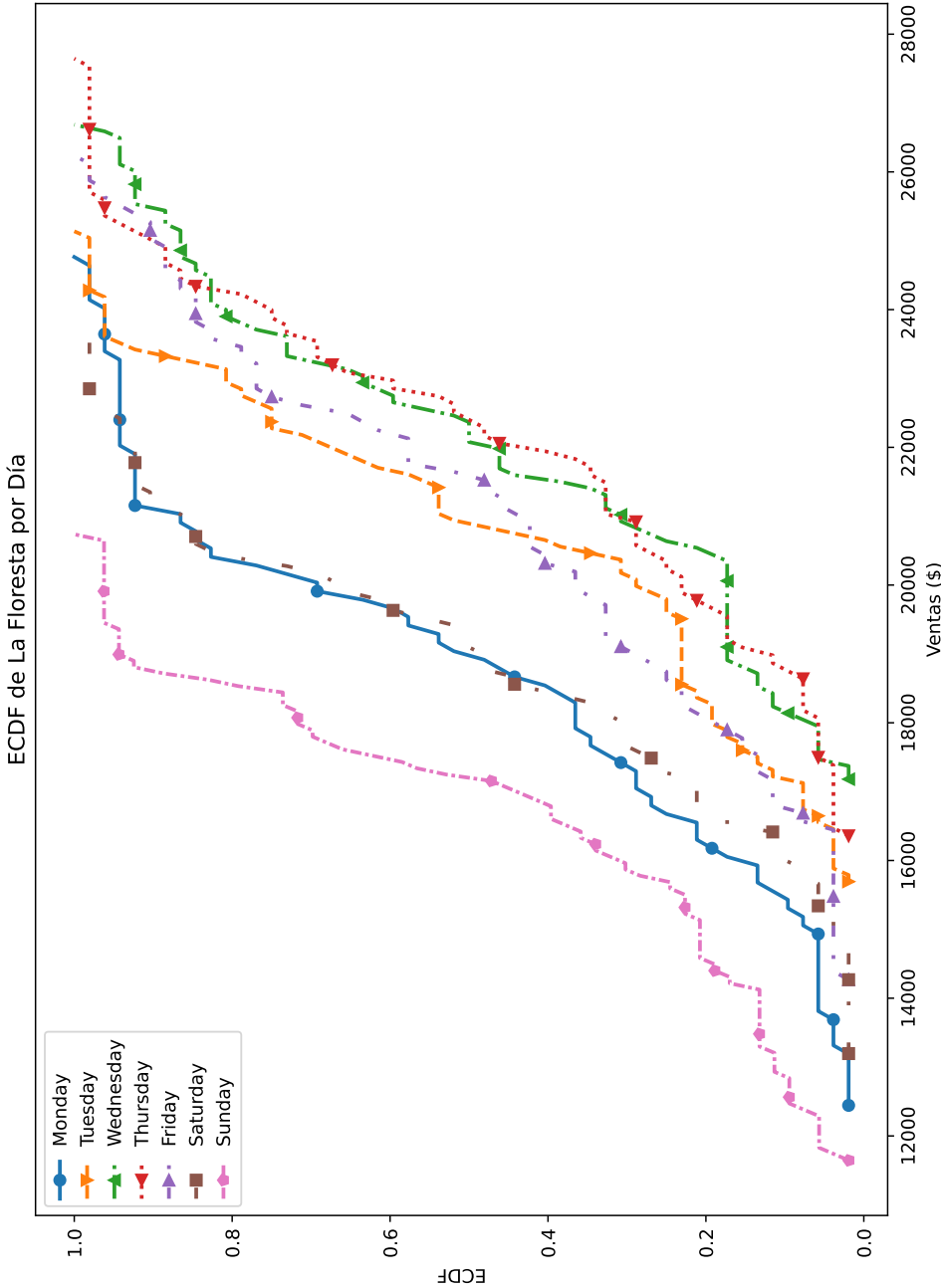
ECDF Santa Ana - DIARIO -



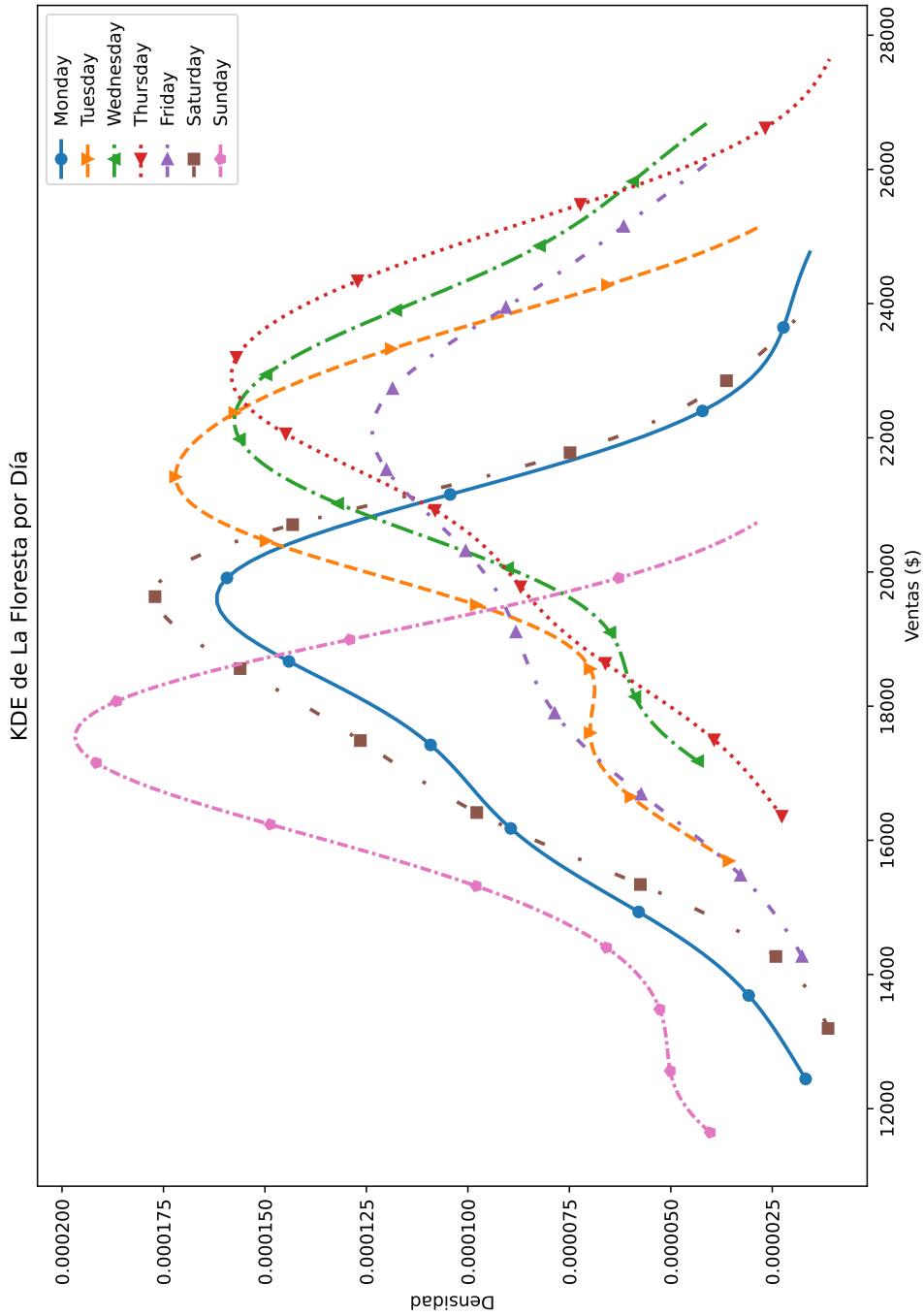
KDE Santa Ana - DIARIO -



ECDF La Floresta - DIARIO -



KDE La Floresta - DIARIO -



IC Santa Ana - Mensual -

Mes	CI 95%	CI 99%
1	8147.54 : 17783.55	7847.39 : 18692.46
2	12245.35 : 19997.85	11520.92 : 21021.57
3	15054.96 : 22967.58	14492.29 : 24044.04
4	11466.17 : 21334.59	10504.82 : 21744.43
5	12646.30 : 21183.46	12500.68 : 21775.03
6	14498.41 : 23340.76	13911.24 : 23431.47
7	11132.45 : 22284.11	11050.94 : 23369.61
8	13256.13 : 22215.41	13211.57 : 22404.25
9	14177.59 : 23413.52	12888.24 : 23587.80
10	14611.63 : 24164.27	14457.83 : 24728.17
11	13747.61 : 22377.59	13532.83 : 23960.70
12	10895.98 : 21487.05	9789.61 : 21985.07

IC La Floresta - Mensual -

Mes	CI 95%	CI 99%
1	11660.34 : 19972.88	11648.15 : 20435.80
2	14434.46 : 22566.85	14250.17 : 23514.12
3	13726.07 : 24095.48	12155.71 : 24267.97
4	15476.25 : 22118.44	15304.44 : 23067.65
5	15806.35 : 24934.21	15333.02 : 25412.73
6	16450.58 : 25987.04	16023.17 : 27313.02
7	13739.18 : 24726.52	12696.44 : 25936.63
8	17477.39 : 26544.22	17456.60 : 26590.06
9	16390.52 : 26253.10	16120.50 : 26598.60
10	17179.83 : 24980.46	17108.24 : 25106.12
11	16021.52 : 25324.06	15767.66 : 25432.83
12	13279.87 : 23071.61	13257.83 : 23729.33

IC Santa Ana - Diario -

Día	CI 95%	CI 99%
Monday	13292.17 : 20693.11	12837.99 : 21390.49
Tuesday	13032.46 : 22697.44	11322.36 : 24345.52
Wednesday	16068.55 : 24128.30	14162.91 : 24345.43
Thursday	15792.53 : 23195.41	13547.84 : 23808.15
Friday	12561.80 : 22500.47	10131.75 : 23123.87
Saturday	12078.41 : 22055.46	8833.12 : 22816.80
Sunday	9547.61 : 19130.55	8595.11 : 19747.99

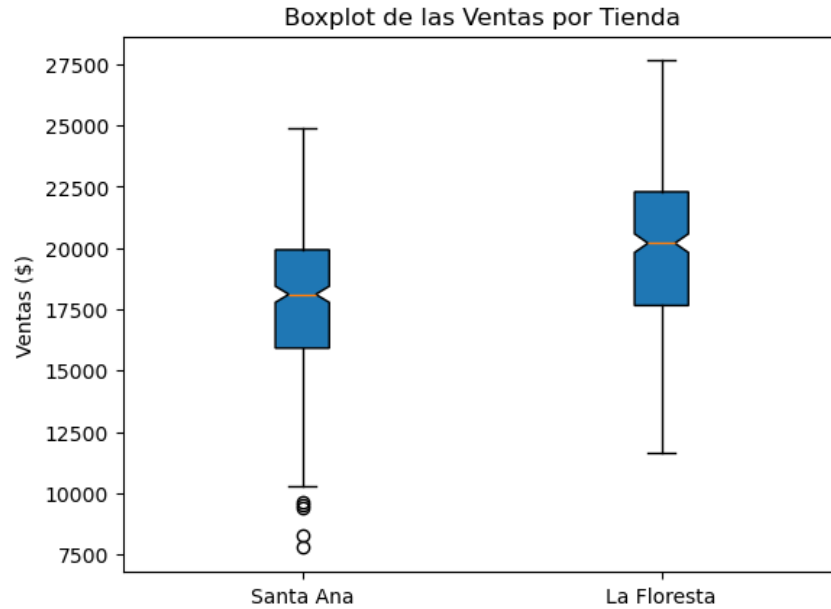
IC La Floresta - Diario -

Día	CI 95%	CI 99%
Monday	13418.38 : 23876.26	12659.25 : 24593.40
Tuesday	16037.55 : 24063.91	15739.89 : 24914.11
Wednesday	17377.56 : 26580.51	17229.88 : 26663.69
Thursday	16671.75 : 25507.71	16364.63 : 27122.38
Friday	14959.57 : 25737.28	14299.46 : 26131.56
Saturday	14948.91 : 22717.67	13612.83 : 23556.20
Sunday	11694.73 : 20271.13	11650.38 : 20714.10

Pruebas de hipotesis

Media de ventas para Santa Ana: 17753.89

Media de ventas para La Floresta: 19987.07



H0: Ventas de ambas tiendas son iguales:

Estadístico t: 9.857108820588982, p-valor: 1.3376077321057801e-21

Rechazamos la hipótesis nula al 95% de confianza (las ventas son diferentes)

Rechazamos la hipótesis nula al 99% de confianza (las ventas son diferentes)

H0: Ventas de 'La Floresta' son mayores o iguales que las de 'Santa Ana'.:

Estadístico t: 9.857108820588982, p-valor: 1.0