

Informe de Análisis de Ventas y Morosidad

Grupo: Rodrigo Goñi

21 de junio de 2025

Planteamiento de los Problemas de Investigación

Don Francisco, propietario de cinco supermercados ('Santa Ana', 'La Floresta', 'Los Cedros', 'Palermo' y 'Córdoba'), ha solicitado a Matías, un especialista en Inteligencia Artificial, un análisis detallado de sus operaciones. Las inquietudes de Don Francisco se centran en tres áreas principales:

1. Entender mejor las ventas mensuales del supermercado 'Santa Ana' mediante el cálculo de intervalos de confianza empíricos.
2. Determinar si las ventas son homogéneas en todos los supermercados o si existen diferencias significativas que requieran atención.
3. Modelar el porcentaje de morosidad de sus clientes a crédito y explicar su comportamiento de pago, utilizando inferencia bayesiana.

Procesamiento Realizado a los Datos

El análisis se realizó utilizando el lenguaje de programación Python y las librerías 'pandas', 'numpy', 'scipy' y 'statsmodels'. Se cargaron los datos de ventas de cada supermercado desde un archivo Excel.

Para el supermercado 'Santa Ana', se extrajeron las ventas diarias, se agruparon por mes y se calcularon la media, la desviación estándar y el número de observaciones por mes. Con estos datos, se construyeron intervalos de confianza empíricos utilizando la distribución t de Student para niveles de significancia del 95 % y 99 %.

Para comparar las ventas entre todos los supermercados, se consolidaron los datos de ventas de todas las hojas del archivo Excel en un único DataFrame. Se realizó un Análisis de Varianza (ANOVA) para determinar si existían diferencias significativas en las medias de ventas entre las tiendas. Posteriormente, en caso de que el ANOVA resultara significativo, se aplicó una prueba post-hoc de Tukey HSD para identificar cuáles pares de supermercados presentaban diferencias estadísticamente significativas, prestando especial atención a la tienda con mayor y menor promedio de ventas.

Resultados Encontrados

1.1. Fundamento Teórico del Intervalo de Confianza

Para este análisis, donde la desviación estándar de la población de ventas es desconocida, se utiliza la distribución t de Student. La fórmula para calcular el intervalo de confianza para la media poblacional (μ) es:

$$IC = \bar{x} \pm t_{\alpha/2, n-1} \times \frac{s}{\sqrt{n}}$$

Donde:

- \bar{x} es la media muestral de las ventas.
- $t_{\alpha/2, n-1}$ es el valor crítico de la distribución t de Student para un nivel de confianza de $1 - \alpha$ y $n - 1$ grados de libertad.
- s es la desviación estándar muestral.
- n es el tamaño de la muestra (número de días de venta).

Este intervalo nos proporciona un rango donde, con un cierto nivel de confianza (e.g., 95 %), podemos esperar que se encuentre la verdadera media de ventas diarias para un mes determinado.

1.2 Intervalos de Confianza Empíricos para 'Santa Ana'

A continuación, se presentan las estadísticas mensuales y los intervalos de confianza calculados para el supermercado 'Santa Ana':

Cuadro 1: Estadísticas Mensuales de Ventas para 'Santa Ana'

AñoMes	Media_Ventas	Desviacion_Estandar	N_Observaciones
2023-01	14674.53	2344.42	31
2023-02	17299.82	2302.73	28
2023-03	18984.29	2670.72	31
2023-04	18721.22	3054.60	30
2023-05	19065.66	2994.60	31
2023-06	19702.58	2412.42	30
2023-07	18778.59	2960.96	31
2023-08	20533.43	2713.14	31
2023-09	20808.33	2627.41	30
2023-10	20394.05	2327.34	31
2023-11	20433.49	2841.50	30
2023-12	18188.54	2213.86	31

- **Mes: 2023-01** Ventas Promedio: \$14,674.53
Nº de Días de Venta: 31
Desviación Estándar (Ventas Diarias): \$2,344.42
IC 95 %: [\$13,814.59, \$15,534.47]
IC 99 %: [\$13,516.59, \$15,832.47]
- **Mes: 2023-02** Ventas Promedio: \$17,299.82
Nº de Días de Venta: 28
Desviación Estándar (Ventas Diarias): \$2,302.73
IC 95 %: [\$16,406.92, \$18,192.73]
IC 99 %: [\$16,094.09, \$18,505.55]
- **Mes: 2023-03** Ventas Promedio: \$18,984.29
Nº de Días de Venta: 31
Desviación Estándar (Ventas Diarias): \$2,670.72
IC 95 %: [\$18,004.66, \$19,963.92]
IC 99 %: [\$17,665.18, \$20,303.39]
- **Mes: 2023-04** Ventas Promedio: \$18,721.22
Nº de Días de Venta: 30
Desviación Estándar (Ventas Diarias): \$3,054.60
IC 95 %: [\$17,580.61, \$19,861.82]
IC 99 %: [\$17,184.01, \$20,258.43]
- **Mes: 2023-05** Ventas Promedio: \$19,065.66
Nº de Días de Venta: 31
Desviación Estándar (Ventas Diarias): \$2,994.60
IC 95 %: [\$17,967.23, \$20,164.08]
IC 99 %: [\$17,586.58, \$20,544.73]
- **Mes: 2023-06** Ventas Promedio: \$19,702.58
Nº de Días de Venta: 30
Desviación Estándar (Ventas Diarias): \$2,412.42
IC 95 %: [\$18,801.76, \$20,603.39]
IC 99 %: [\$18,488.54, \$20,916.62]
- **Mes: 2023-07** Ventas Promedio: \$18,778.59
Nº de Días de Venta: 31
Desviación Estándar (Ventas Diarias): \$2,960.96
IC 95 %: [\$17,692.50, \$19,864.68]
IC 99 %: [\$17,316.13, \$20,241.05]
- **Mes: 2023-08** Ventas Promedio: \$20,533.43
Nº de Días de Venta: 31
Desviación Estándar (Ventas Diarias): \$2,713.14
IC 95 %: [\$19,538.24, \$21,528.61]
IC 99 %: [\$19,193.37, \$21,873.48]

- **Mes: 2023-09** Ventas Promedio: \$20,808.33
 N° de Días de Venta: 30
 Desviación Estándar (Ventas Diarias): \$2,627.41
 IC 95 %: [\$19,827.24, \$21,789.42]
 IC 99 %: [\$19,486.10, \$22,130.56]
- **Mes: 2023-10** Ventas Promedio: \$20,394.05
 N° de Días de Venta: 31
 Desviación Estándar (Ventas Diarias): \$2,327.34
 IC 95 %: [\$19,540.37, \$21,247.72]
 IC 99 %: [\$19,244.54, \$21,543.55]
- **Mes: 2023-11** Ventas Promedio: \$20,433.49
 N° de Días de Venta: 30
 Desviación Estándar (Ventas Diarias): \$2,841.50
 IC 95 %: [\$19,372.46, \$21,494.52]
 IC 99 %: [\$19,003.52, \$21,863.46]
- **Mes: 2023-12** Ventas Promedio: \$18,188.54
 N° de Días de Venta: 31
 Desviación Estándar (Ventas Diarias): \$2,213.86
 IC 95 %: [\$17,376.49, \$19,000.59]
 IC 99 %: [\$17,095.08, \$19,281.99]

2. Análisis de Varianza (ANOVA) para Comparar Ventas de Todas las Tiendas

2.1. Fundamento Teórico del Análisis de Varianza (ANOVA)

El Análisis de Varianza (ANOVA) es una prueba estadística diseñada para comparar las medias de tres o más grupos y determinar si al menos una de ellas es diferente de las demás. En este caso, se utiliza para comparar la media de ventas diarias entre los cinco supermercados.

La prueba ANOVA se basa en el **estadístico F**, que compara la variabilidad *entre* los grupos con la variabilidad *dentro* de los grupos. Si la variabilidad entre los grupos es significativamente mayor que la variabilidad dentro de los grupos, se concluye que no todas las medias son iguales.

El estadístico F se calcula con la siguiente fórmula:

$$F = \frac{\text{Varianza entre grupos}}{\text{Varianza dentro de los grupos}} = \frac{\sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 / (k - 1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / (N - k)}$$

Donde:

- k es el número de grupos (supermercados).
- n_i es el número de observaciones en el grupo i .

- N es el número total de observaciones.
- \bar{x}_i es la media del grupo i .
- \bar{x} es la media total de todas las observaciones.
- x_{ij} es la observación j del grupo i .

Un valor F grande, asociado a un p -valor pequeño (generalmente ≤ 0.05), nos lleva a rechazar la hipótesis nula de que todas las medias son iguales.

2.2. Verificación de Supuestos del ANOVA

Para determinar si existen diferencias significativas en el rendimiento de ventas entre los distintos supermercados, se procedió a realizar un Análisis de Varianza (ANOVA). Sin embargo, para garantizar la validez de los resultados de esta prueba, primero es fundamental verificar que se cumplan sus supuestos estadísticos. a) Independencia: Se asume que las observaciones son independientes por el diseño del estudio. Antes de interpretar los resultados del ANOVA, se evaluaron dos supuestos clave: la homogeneidad de varianzas y la normalidad de los residuos.

- **Homogeneidad de Varianzas (Prueba de Levene):** Este supuesto requiere que la dispersión de los datos (varianza) sea similar en todos los supermercados.
 - **Hipótesis Nula (H_0):** Las varianzas de ventas son iguales en todos los grupos.
 - **Resultado:** La prueba arrojó un p -valor de **0.4965**. Al ser este valor considerablemente mayor que el nivel de significancia ($\alpha = 0,05$), no se rechaza la hipótesis nula.
 - **Conclusión:** Se cumple el supuesto de homogeneidad de varianzas.
- **Normalidad de los Residuos (Prueba de Shapiro-Wilk):** Este supuesto requiere que los residuos del modelo sigan una distribución normal (curva de campana).
 - **Hipótesis Nula (H_0):** Los residuos se distribuyen de forma normal.
 - **Resultado:** La prueba arrojó un p -valor de **0.0000**. Este valor es inferior al nivel de significancia ($\alpha = 0,05$), por lo que se rechaza la hipótesis nula.
 - **Conclusión:** No se cumple el supuesto de normalidad.

El incumplimiento del supuesto de normalidad, que se confirma visualmente en las Figuras 1 y 2 por la asimetría y la presencia de valores atípicos, indica que los resultados del ANOVA estándar deben ser interpretados con precaución.

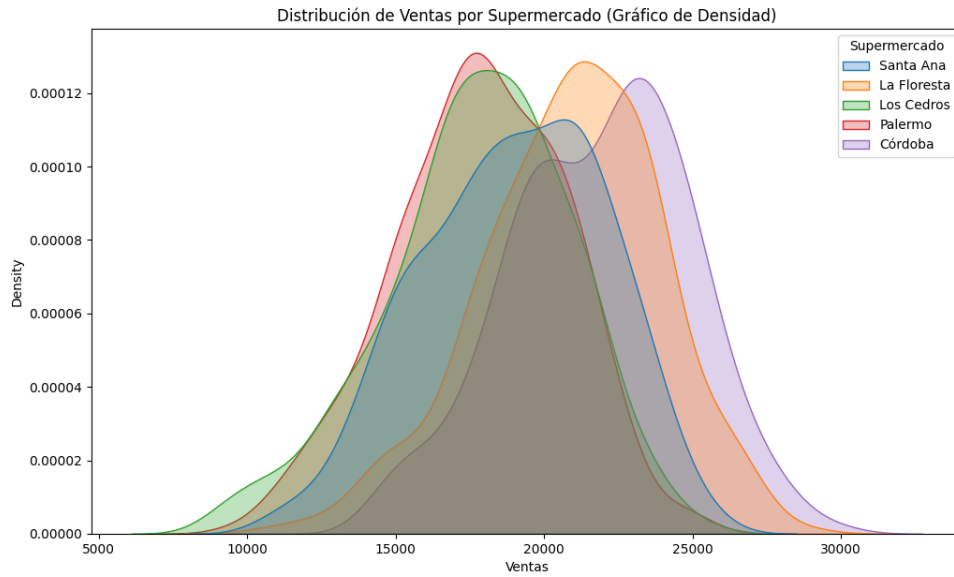


Figura 1: Gráfico de densidad que muestra la distribución de las ventas para cada supermercado. Se observa una forma no perfectamente simétrica en varias de las curvas.

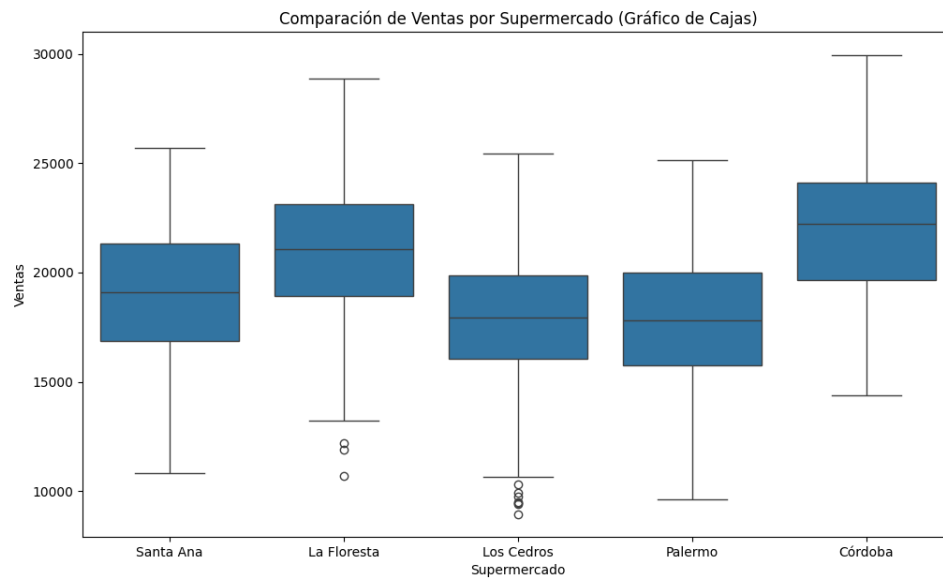


Figura 2: Gráfico de cajas que compara la mediana y la dispersión de las ventas. Se aprecian valores atípicos (outliers) en “La Floresta” y “Los Cedros”, lo que contribuye al incumplimiento del supuesto de normalidad.

2.3. Resultados de la Prueba ANOVA

A pesar de la advertencia sobre la normalidad, se presenta la tabla del ANOVA, ya que sigue siendo informativa, especialmente con un tamaño de muestra grande.

Hipótesis del Análisis:

- **Hipótesis Nula (H_0):** Las ventas promedio diarias de todos los supermercados son iguales.
- **Hipótesis Alternativa (H_1):** Al menos un supermercado tiene un promedio de ventas diario significativamente diferente.

Cuadro 2: Tabla de Resultados del Análisis de Varianza (ANOVA)

	sum_sq	df	F	PR(>F)
C(Supermercado)	4,916848e + 09	4,0	133,638261	3,511300e - 100
Residual	1,674046e + 10	1820,0	NaN	NaN

El resultado de la prueba muestra un valor F de 133.64 y un p-valor ($PR(>F)$) extremadamente pequeño ($3,51e - 100$).

2.4. Conclusión del Análisis

Con un p-valor prácticamente nulo, **se rechaza la hipótesis nula (H_0) de forma contundente**. Existe evidencia estadística abrumadora para afirmar que **no todas las tiendas se comportan de la misma manera en términos de ventas**.

Dado que el supuesto de normalidad no se cumplió, la aproximación más robusta sería confirmar este hallazgo con una **prueba no paramétrica como la de Kruskal-Wallis**. No obstante, la magnitud del resultado del ANOVA, junto con la clara evidencia visual de los gráficos, nos da una alta confianza para proceder al siguiente paso: realizar un **análisis post-hoc** (como la prueba de Tukey HSD) para identificar exactamente cuáles son los supermercados que presentan diferencias significativas entre sí.

3. Comparación de la Tienda con Mayor y Menor Promedio de Ventas

3.1. Fundamento Teórico de la Prueba de Tukey HSD

Cuando una prueba ANOVA resulta significativa (como en el paso anterior), nos informa que existe una diferencia en las medias de los grupos, pero no especifica entre qué pares de grupos se encuentra esa diferencia. Para resolver esta cuestión, se emplean las ****pruebas post-hoc****.

La prueba de ****Diferencia Honestamente Significativa de Tukey (Tukey HSD)**** es uno de los métodos post-hoc más robustos y utilizados. Su función principal es realizar comparaciones por pares de todas las medias de los grupos (por ejemplo, comparar Córdoba vs. Los Cedros, Córdoba vs. Palermo, etc.) mientras controla la ****tasa de error familiar****. Esto es crucial porque al realizar múltiples comparaciones, la probabilidad de cometer un error de Tipo I (un falso positivo) aumenta; la prueba de Tukey ajusta los p-valores para mantener esta tasa de error general en el nivel deseado (usualmente 5

La prueba calcula un estadístico llamado ****rango estudentizado (q)**** para cada par de medias. La fórmula es:

$$q = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{\frac{MS_{dentro}}{n_h}}}$$

Donde:

- \bar{x}_i y \bar{x}_j son las medias de los dos grupos que se están comparando.
- MS_{dentro} (Mean Square within-group) es la varianza promedio dentro de los grupos, que se obtiene directamente de la tabla ANOVA (es la suma de cuadrados residual, $sumsq$ Residual, dividida por los grados de libertad residuales, df Residual). Este valor representa la variabilidad aleatoria o no explicada.
- n_h es el tamaño de la muestra armónica, que se utiliza cuando los tamaños de los grupos son desiguales. Si son iguales, $n_h = n$ (el tamaño de la muestra por grupo).

El valor q calculado se compara con un valor crítico de la distribución del rango estudentizado. Esto genera un ****p-valor ajustado ('p-adj')**** para cada par. Si este 'p-adj' es menor que el nivel de significancia ($\alpha = 0,05$), se rechaza la hipótesis nula para ese par y se concluye que la diferencia entre sus medias es estadísticamente significativa.

3.2. Aplicación y Resultados: Tiendas con Mayor y Menor Promedio

Primero, identificamos los promedios de ventas para ordenar las tiendas de mayor a menor rendimiento: Los promedios de ventas por supermercado son los siguientes:

- Córdoba: \$21,868.46
- La Floresta: \$20,874.02
- Santa Ana: \$18,968.64
- Palermo: \$17,835.95
- Los Cedros: \$17,789.07

La tienda con mayor promedio de ventas es 'Córdoba' (\$21,868.46) y la tienda con menor promedio de ventas es 'Los Cedros' (\$17,789.07).

Se realizó una prueba de Tukey HSD (Post-hoc ANOVA) para comparaciones pareadas, dado que la prueba ANOVA inicial resultó significativa.

Para la comparación específica entre 'Córdoba' (mayor promedio) y 'Los Cedros' (menor promedio):

- Diferencia de medias: \$-4,079.38
- P-value (Tukey HSD): 0.0000

La diferencia entre las ventas de la tienda con mayor y menor promedio es estadísticamente significativa. Esto sugiere que 'Los Cedros' realmente tiene ventas peores y podría necesitar más atención.

Cuadro 3: Resultados de la Prueba Tukey HSD

group1	group2	meandiff	p-adj	lower	upper	reject
Córdoba	La Floresta	-994.4327	0.0001	-1607.433	-381.4323	True
Córdoba	Los Cedros	-4079.3818	0.0	-4692.3822	-3466.3815	True
Córdoba	Palermo	-4032.5033	0.0	-4645.5037	-3419.503	True
Córdoba	Santa Ana	-2899.8136	0.0	-3512.814	-2286.8133	True
La Floresta	Los Cedros	-3084.9492	0.0	-3697.9495	-2471.9488	True
La Floresta	Palermo	-3038.0707	0.0	-3651.071	-2425.0703	True
La Floresta	Santa Ana	-1905.381	0.0	-2518.3813	-1292.3806	True
Los Cedros	Palermo	46.8785	0.9996	-566.1218	659.8789	False
Los Cedros	Santa Ana	1179.5682	0.0	566.5679	1792.5686	True
Palermo	Santa Ana	1132.6897	0.0	519.6894	1745.69	True

Análisis y Conclusiones

1. **Ventas de 'Santa Ana':** Los intervalos de confianza empíricos para cada mes proporcionan a Don Francisco un rango estimado de ventas con un 95 % y 99 % de certeza. Esto es crucial para la planificación financiera y la identificación de patrones estacionales. Por ejemplo, se observa que las ventas en enero de 2023 fueron las más bajas, con un promedio de \$14,674.53, mientras que los meses de agosto, septiembre y octubre mostraron promedios superiores a \$20,000.
2. **Comparación entre Supermercados (ANOVA):** La prueba ANOVA reveló que existen diferencias estadísticamente significativas en las ventas promedio entre los cinco supermercados. Esto significa que no todas las tiendas tienen el mismo rendimiento de ventas y, por lo tanto, la estrategia de ventas o la asignación de recursos podría necesitar ser diferenciada.
3. **Tiendas con Mayor y Menor Rendimiento:** El supermercado 'Córdoba' se identificó como el de mayor promedio de ventas, mientras que 'Los Cedros' tuvo el menor promedio. La prueba de Tukey HSD confirmó que la diferencia de ventas entre 'Córdoba' y 'Los Cedros' es estadísticamente significativa. Esto valida la inquietud de Don Francisco y sugiere que 'Los Cedros' podría requerir una atención especial, como campañas de marketing focalizadas, mejoras en la gestión o una revisión de su oferta de productos, para mejorar su rendimiento.

Anexo: Código Python

```
1 import numpy as np
2 from scipy import stats
3 import statsmodels.api as sm
4 from statsmodels.formula.api import ols
5 from statsmodels.stats.multicomp import pairwise_tukeyhsd
6
7
8 def analyze_sales_data(file_path, santa_ana_sheet_name='Santa Ana'):
9     """
10     Realiza los analisis de ventas para los supermercados.
11
12     Args:
13         file_path (str): Ruta al archivo XLSX.
14         santa_ana_sheet_name (str): Nombre de la hoja del supermercado 'Santa
15             Ana'.
16     """
17     xls = pd.ExcelFile(file_path)
18     sheet_names = xls.sheet_names
19
20     # 1. Intervalos de Confianza Empiricos para 'Santa Ana'
21     print("--- 1. Intervalos de Confianza Empiricos para 'Santa Ana' ---")
22     if santa_ana_sheet_name in sheet_names:
23         df_santa_ana = pd.read_excel(xls, sheet_name=santa_ana_sheet_name)
24         df_santa_ana.columns = ['Fecha', 'Ventas']
25         df_santa_ana['Fecha'] = pd.to_datetime(
26             df_santa_ana['Fecha']) # Convertir a datetime
27
28         df_santa_ana['AnoMes'] = df_santa_ana['Fecha'].dt.to_period('M')
29
30         monthly_sales = df_santa_ana.groupby('AnoMes')['Ventas'].agg(
31             ['mean', 'std', 'count']).reset_index()
32         monthly_sales.rename(columns={
33             'mean': 'Media_Ventas', 'std': 'Desviacion_Estandar',
34             'count': 'N_Observaciones'}, inplace=True)
35
36         print("\nEstadísticas mensuales de ventas (agrupadas por mes):")
37         print(monthly_sales)
38         print("-" * 50) # Separador para mejor legibilidad
39
40         for index, row in monthly_sales.iterrows():
41             mes_ano = row['AnoMes']
42             media_ventas = row['Media_Ventas']
43             desviacion_estandar = row['Desviacion_Estandar']
44             n_observaciones = row['N_Observaciones']
```

```

44     if n_observaciones > 1 and not pd.isna(desviacion_estandar):
45         se = desviacion_estandar / np.sqrt(n_observaciones)
46
47         t_crit_95 = stats.t.ppf(0.975, n_observaciones - 1)
48         lower_95 = media_ventas - t_crit_95 * se
49         upper_95 = media_ventas + t_crit_95 * se
50
51         t_crit_99 = stats.t.ppf(0.995, n_observaciones - 1)
52         lower_99 = media_ventas - t_crit_99 * se
53         upper_99 = media_ventas + t_crit_99 * se
54
55         print(f"Mes: {mes_ano}")
56         print(f" Ventas Promedio: ${media_ventas:,.2f}")
57         print(f" N de Dias de Venta: {n_observaciones}")
58         print(
59             f" Desviacion Estandar (Ventas Diarias):
60                 ${desviacion_estandar:,.2f}")
61         print(f" IC 95%: [{lower_95:,.2f}, {upper_95:,.2f}]")
62         print(f" IC 99%: [{lower_99:,.2f}, {upper_99:,.2f}]")
63         print("-" * 30)
64     else:
65         print(f"Mes: {mes_ano} - No hay suficientes datos
66             ({n_observaciones} observaciones) o la desviacion estandar es
67             nula para calcular el IC. Se necesitan al menos 2 dias de
68             venta por mes.")
69         print("-" * 30)
70     print("\n")
71 else:
72     print(
73         f"La hoja '{santa_ana_sheet_name}' no se encontro en el archivo.")
74     print("\n")
75
76 # 2. Pruebas ANOVA para determinar si las ventas esperadas de todas las
77     tiendas son iguales
78 print("--- 2. Pruebas ANOVA para Comparar Ventas de Todas las Tiendas ---")
79 all_sales_data = []
80 for sheet_name in sheet_names:
81     df_temp = pd.read_excel(xls, sheet_name=sheet_name)
82     df_temp.columns = ['Fecha', 'Ventas']
83     df_temp['Supermercado'] = sheet_name
84     all_sales_data.append(df_temp)
85
86 df_all_sales = pd.concat(all_sales_data, ignore_index=True)
87
88 if len(df_all_sales['Supermercado'].unique()) < 2:
89     print("Se necesitan al menos dos supermercados para realizar ANOVA.")
90 elif df_all_sales['Ventas'].isnull().any():

```

```

86     print("Advertencia: Se encontraron valores nulos en las ventas.
      Eliminandolos para ANOVA.")
87     df_all_sales.dropna(subset=['Ventas'], inplace=True)
88     if df_all_sales.empty:
89         print(
90             "No quedan datos despues de eliminar los nulos. No se puede
              realizar ANOVA.")
91         return
92
93     # Realizar la prueba ANOVA
94     model = ols('Ventas ~ C(Supermercado)', data=df_all_sales).fit()
95     anova_table = sm.stats.anova_lm(model, typ=2)
96
97     print("Tabla ANOVA:")
98     print(anova_table)
99
100    alpha = 0.05
101    if 'C(Supermercado)' in anova_table.index:
102        p_value_anova = anova_table.loc['C(Supermercado)', 'PR(>F)']
103    else:
104
105        p_value_anova = anova_table['PR(>F)'][0]
106
107    if p_value_anova < alpha:
108        print(
109            f"\nCon un nivel de significancia del {alpha*100}%, rechazamos la
              hipotesis nula.")
110        print("Hay evidencia estadistica de que las ventas esperadas de al menos
              un supermercado son diferentes.")
111        print("Esto sugiere que no todas las tiendas se comportan de la misma
              manera en terminos de ventas.")
112    else:
113        print(
114            f"\nCon un nivel de significancia del {alpha*100}%, no rechazamos la
              hipotesis nula.")
115        print("No hay evidencia estadistica de que las ventas esperadas de los
              supermercados sean diferentes.")
116        print("Las ventas promedio de los supermercados son estadisticamente
              similares.")
117    print("\n")
118
119    # 3. Identificar la tienda con mayor y menor promedio de ventas y prueba de
      hipotesis
120    print("--- 3. Comparacion de la Tienda con Mayor y Menor Promedio de Ventas
      ---")
121
122    if not df_all_sales.empty:

```

```

123 avg_sales_by_supermarket = df_all_sales.groupby(
124     'Supermercado')['Ventas'].mean().sort_values(ascending=False)
125 print("Promedio de ventas por supermercado:")
126 print(avg_sales_by_supermarket)
127
128 supermarket_highest_sales = avg_sales_by_supermarket.index[0]
129 supermarket_lowest_sales = avg_sales_by_supermarket.index[-1]
130
131 print(
132     f"\nTienda con mayor promedio de ventas:
133         '{supermarket_highest_sales}'
134         (${avg_sales_by_supermarket.iloc[0]:,.2f})")
135
136 print(
137     f"Tienda con menor promedio de ventas: '{supermarket_lowest_sales}'
138         (${avg_sales_by_supermarket.iloc[-1]:,.2f})")
139
140 if p_value_anova < alpha:
141     print(
142         "\nRealizando prueba de Tukey HSD (Post-hoc ANOVA) para
143             comparaciones pareadas:")
144     tukey_result = pairwise_tukeyhsd(endog=df_all_sales['Ventas'],
145                                     groups=df_all_sales['Supermercado'],
146                                     alpha=0.05)
147     print(tukey_result)
148
149     tukey_df = pd.DataFrame(data=tukey_result._results_table.data[1:],
150                             columns=tukey_result._results_table.data[0])
151
152     found_comparison = False
153     specific_comparison = tukey_df[
154         ((tukey_df['group1'] == supermarket_highest_sales) &
155          (tukey_df['group2'] == supermarket_lowest_sales)) |
156         ((tukey_df['group1'] == supermarket_lowest_sales) &
157          (tukey_df['group2'] == supermarket_highest_sales))
158     ]
159
160     if not specific_comparison.empty:
161         p_value_tukey = specific_comparison['p-adj'].iloc[0]
162         mean_diff_tukey = specific_comparison['meandiff'].iloc[0]
163
164         print(
165             f"\nComparacion especifica entre '{supermarket_highest_sales}'
166                 y '{supermarket_lowest_sales}':")
167         if specific_comparison['group1'].iloc[0] ==
168             supermarket_lowest_sales:
169             mean_diff_tukey *= -1
170         print(f" Diferencia de medias: ${mean_diff_tukey:,.2f}")

```

```

163     print(f" P-value (Tukey HSD): {p_value_tukey:.4f}")
164     found_comparison = True
165
166     if p_value_tukey < alpha:
167         print(
168             "La diferencia entre las ventas de la tienda con mayor y
169             menor promedio es estadísticamente significativa.")
170         print(
171             f"Esto sugiere que '{supermarket_lowest_sales}' realmente
172             tiene ventas peores y podría necesitar más atención.")
173     else:
174         print(
175             "La diferencia entre las ventas de la tienda con mayor y
176             menor promedio NO es estadísticamente significativa.")
177         print("Las diferencias observadas podrían ser debidas al
178             azar.")
179
180     if not found_comparison:
181         print(
182             f"No se encontró una comparación directa entre
183             '{supermarket_highest_sales}' y
184             '{supermarket_lowest_sales}' en los resultados de Tukey
185             HSD.")
186         print(
187             "Esto es inusual si ambas tiendas están en el dataset.
188             Verifique los nombres.")
189     else:
190         print("\nEl ANOVA no encontró diferencias significativas, por lo que
191             una prueba post-hoc directa entre las tiendas con mayor y menor
192             promedio no es estadísticamente necesaria para confirmar una
193             diferencia global. Las diferencias observadas son probablemente
194             debidas al azar.")
195
196     else:
197         print("No hay datos disponibles para realizar la comparación de tiendas.")
198     print("\n")
199
200 if __name__ == "__main__":
201     excel_file_path = "Datos_examen_final_21Co20258_a2119.xlsx"
202     analyze_sales_data(excel_file_path)

```

Listing 1: Código Python