Análisis del Impacto del Consumo de Alcohol en Accidentes de Tránsito durante Fiestas Patrias

Nombre: Roger Fuentes

Asignatura: minería de datos

Fecha: 06/10/2025

Contenido

[Ejercicio 1 3](#_Toc210489458)

[Ejercicio 2 4](#_Toc210489459)

[Ejercicio 3 5](#_Toc210489460)

[Ejercicio 4 6](#_Toc210489461)

[Ejercicio 5 9](#_Toc210489462)

[Ejercicio 6 11](#_Toc210489463)

[Ejercicio 7 13](#_Toc210489464)

[Ejercicio 8 15](#_Toc210489465)

# Ejercicio 1

Muestra las primeras 10 filas del dataset\_fiestas\_patrias.csv.

* ¿Cuántas filas y columnas tiene el dataset?
* ¿Cuántas regiones distintas aparecen en la columna región?

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

df = pd.read\_csv("fiestas.csv")

print("Primeras 10 filas del DataFrame:")

print(df.head(**10**))

# Obtener y mostrar el numero de filas y columnas

filas, columnas = df.shape

print(f"**\n**El dataset tiene {filas} filas y {columnas} columnas.")

# Obtener y mostrar el numero de regiones distintas

regiones\_distintas = df['region'].nunique()

print(f"Aparecen {regiones\_distintas} regiones distintas en la columna 'region'.")

Texto

El contenido generado por IA puede ser incorrecto.

# Ejercicio 2

En el ejercicio 2 se nos pide:

* Calcular el promedio y la mediana de litros\_bebida\_alcoholica.
* Calcular el máximo, mínimo y promedio de kg\_carne
* ¿Cuál es la edad promedio de los asistentes a nivel general?

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

df = pd.read\_csv("fiestas.csv")

promedio\_alcohol = df['litros\_bebida\_alcoholica'].mean()

mediana\_alcohol = df['litros\_bebida\_alcoholica'].median()

print(f"**\n**El promedio de 'litros\_bebida\_alcoholica' es: {promedio\_alcohol:.2f}")

print(f"La mediana de 'litros\_bebida\_alcoholica' es: {mediana\_alcohol:.2f}")

max\_carne = df['kg\_carne'].max()

min\_carne = df['kg\_carne'].min()

promedio\_carne = df['kg\_carne'].mean()

print(f"**\n**El máximo de 'kg\_carne' es: {max\_carne}")

print(f"El mínimo de 'kg\_carne' es: {min\_carne}")

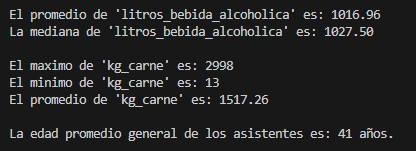
print(f"El promedio de 'kg\_carne' es: {promedio\_carne:.2f}")

# Edad promedio general de los asistentes

promedio\_edad\_general = df['edad\_promedio\_asistentes'].mean()

print(f"**\n**La edad promedio general de los asistentes es: {promedio\_edad\_general:.0f} años.")

print("**\n**")



# Ejercicio 3

Datos solicitados.

* Agrupar el dataset por región y calcular el total de fallecidos y heridos.
* ¿Qué región representa más detenidos en total?
* ¿Qué tipo de evento (tipo\_evento) tiene en promedio más asistentes?

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

df = pd.read\_csv("fiestas.csv")

fallecidos\_heridos\_por\_region = df.groupby('region')[['fallecidos', 'heridos']].sum()

print("**\n**1. Total de fallecidos y heridos por región:")

print(fallecidos\_heridos\_por\_region)

detenidos\_por\_region = df.groupby('region')['detenidos'].sum()

region\_mas\_detenidos = detenidos\_por\_region.idxmax()

total\_detenidos = detenidos\_por\_region.max()

print(f"**\n**2. La región con más detenidos en total es '{region\_mas\_detenidos}' con {total\_detenidos} detenidos.")

asistentes\_por\_evento = df.groupby('tipo\_evento')['n\_asistentes'].mean()

evento\_mas\_asistentes = asistentes\_por\_evento.idxmax()

promedio\_asistentes = asistentes\_por\_evento.max()

print(f"**\n**3. El tipo de evento con más asistentes en promedio es '{evento\_mas\_asistentes}' con {promedio\_asistentes:.0f} asistentes.")

print("**\n**")

Pantalla de computadora con letras

El contenido generado por IA puede ser incorrecto.

# Ejercicio 4

Datos solicitados:

* Seleccionar los eventos con más de **2000 asistentes.**
* Seleccionar los eventos donde hubo mas de **500 litros de alcohol** y al menos **1 fallecido**
* ¿Cuántos eventos de estas características se realizaron en condiciones lluvioso?

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

df = pd.read\_csv("fiestas.csv")

# eventos con mas de 2000 asistentes

eventos\_masivos = df[df['n\_asistentes'] > **2000**]

print("1. Eventos con más de 2000 asistentes:")

print(f"(Mostrando las primeras 5 de {len(eventos\_masivos)} filas encontradas)**\n**", eventos\_masivos.head())

# eventos con mass de 500 litros de alcohol y al menos 1 fallecido

eventos\_criticos = df[(df['litros\_bebida\_alcoholica'] > **500**) & (df['fallecidos'] >= **1**)]

print("**\n\n**2. Eventos con >500L de alcohol y al menos 1 fallecido:")

print(eventos\_criticos)

# eventos anteriores en condiciones de lluvia

eventos\_criticos\_lluviosos = eventos\_criticos[eventos\_criticos['condiciones\_climaticas'] == 'Lluvioso']

cantidad\_lluviosos = len(eventos\_criticos\_lluviosos)

print(f"**\n\n**3. De esos eventos críticos, {cantidad\_lluviosos} se realizaron en condiciones de 'Lluvioso'.")

print("**\n**")

Texto

El contenido generado por IA puede ser incorrecto.

Calendario

El contenido generado por IA puede ser incorrecto.

Texto

El contenido generado por IA puede ser incorrecto.

* Se encontraron 57 eventos críticos, y de esos eventos críticos, 14 se realizaron en condiciones lluviosas.

# Ejercicio 5

Solicitado.

* Calcular la correlación entre litros\_bebida\_alcoholica y accidentes \_transito.
* ¿existe relación entre n\_asistentes y kg\_carne consumidos?

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

df = pd.read\_csv("fiestas.csv")

print("**\n**")

corr\_alcohol\_accidentes = df['litros\_bebida\_alcoholica'].corr(df['accidentes\_transito'])

print("1. Correlación entre 'litros\_bebida\_alcoholica' y 'accidentes\_transito':")

print(f"{corr\_alcohol\_accidentes:.4f}")

corr\_asistentes\_carne = df['n\_asistentes'].corr(df['kg\_carne'])

print("**\n**2. ¿Existe relación entre 'n\_asistentes' y 'kg\_carne' consumidos?")

print(f"Correlación: {corr\_asistentes\_carne:.4f}")

print("**\n**")

Texto

El contenido generado por IA puede ser incorrecto.

* **Correlación entre 'litros\_bebida\_alcoholica' y 'accidentes\_transito': 0.0400**

Este valor indica que la relación lineal entre la cantidad de litros de bebida alcohólica consumidos y la cantidad de accidentes de tránsito es muy baja y positiva. Un valor de correlación cercano a 0 (como 0.0400) significa que prácticamente no existe una relación lineal entre estas dos variables en los datos analizados.

* **¿Existe relación entre 'n\_asistentes' y 'kg\_carne' consumidos? Correlación: -0.0121**

El valor de correlación es -0.0121, lo que también está muy cerca de 0. Esto indica que no hay una relación lineal significativa entre el número de asistentes y los kilogramos de carne consumidos en los eventos. El signo negativo muestra una tendencia muy ligera a que, si una variable aumenta, la otra disminuye, pero la relación es tan débil que puede considerarse inexistente.

# Ejercicio 6

* Se solicita graficar un **HISTOGRAMA** del consumo de kg\_carne.
* Hacer un gráfico de barras que muestre la cantidad de eventos por región.

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**import** **matplotlib.pyplot** **as** **plt**

df = pd.read\_csv("fiestas.csv")

plt.figure(figsize=(**10**, **6**))

plt.hist(df['kg\_carne'], bins=**30**, color='skyblue', edgecolor='black')

plt.title('Distribución del Consumo de Carne (kg) en Eventos')

plt.xlabel('Kg de Carne Consumidos')

plt.ylabel('Frecuencia (Número de Eventos)')

plt.grid(axis='y', alpha=**0.75**)

plt.show()

Gráfico, Gráfico de barras, Histograma

El contenido generado por IA puede ser incorrecto.

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**import** **matplotlib.pyplot** **as** **plt**

df = pd.read\_csv("fiestas.csv")

eventos\_por\_region = df['region'].value\_counts()

plt.figure(figsize=(**12**, **7**))

eventos\_por\_region.plot(kind='bar', color='coral', edgecolor='black')

plt.title('Cantidad de Eventos por Región', fontsize=**16**)

plt.xlabel('Región', fontsize=**12**)

plt.ylabel('Número de Eventos', fontsize=**12**)

plt.xticks(rotation=**45**, ha='right')

plt.grid(axis='y', linestyle='--', alpha=**0.7**)

plt.tight\_layout()

plt.show()

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

# Ejercicio 7

* Graficar un **diagrama de dispersión (scatter plot)** entre litros\_bebida\_alcoholica y accidentes\_transito.
* Colorear los puntos según la condiciones\_climaticas.
* ¿Qué observaciones se pueden sacar del grafico?

Codigo

**import** **pandas** **as** **pd**

**import** **matplotlib.pyplot** **as** **plt**

# Cargar el archivo CSV

df = pd.read\_csv("fiestas.csv")

# colores

colores = {

'Soleado': 'gold',

'Nublado': 'grey',

'Lluvioso': 'dodgerblue'

}

plt.figure(figsize=(**8**, **6**))

**for** clima, color **in** colores.items():

datos = df[df['condiciones\_climaticas'] == clima]

plt.scatter(

datos['litros\_bebida\_alcoholica'],

datos['accidentes\_transito'],

color=color,

label=clima,

alpha=**0.7**,

edgecolors='black'

)

plt.title("Alcohol vs Accidentes según Clima")

plt.xlabel("Litros de Bebida Alcohólica")

plt.ylabel("Número de Accidentes de Tránsito")

plt.grid(**True**)

plt.legend(title="Condición Climática")

plt.tight\_layout()

plt.show()

Gráfico

El contenido generado por IA puede ser incorrecto.

El gráfico de dispersión muestra la relación entre los litros de bebida alcohólica consumidos y el número de accidentes de tránsito registrados durante eventos, diferenciados por condición climática.

**Principales hallazgos:**

No se observa una correlación directa entre el consumo de alcohol y la cantidad de accidentes: eventos con alto consumo no siempre presentan más incidentes.

Eventos en clima soleado tienden a tener mayor consumo de alcohol, pero también registran accidentes, lo que podría reflejar mayor actividad social o menor percepción de riesgo.

Eventos en clima lluvioso, aunque con menor consumo, presentan algunos casos con más accidentes, lo que sugiere que las condiciones climáticas adversas podrían aumentar el riesgo vial.

Eventos nublados muestran una distribución más dispersa, con algunos casos de alto consumo y pocos accidentes, lo que podría indicar mejor control o menor exposición.

# Ejercicio 8

**Análisis completo del impacto del consumo en accidentes de Fiestas Patrias**

**¿Existe una relación significativa entre el consumo de alcohol y el número de accidentes de tránsito durante Fiestas Patrias? Además, ¿cómo varía este fenómeno según la región del país?**

**Tareas que deben realizar**

**1. Carga de datos (Pandas):**

* Leer el dataset dataset\_fiestas\_patrias.csv.
* Revisar la estructura del dataset y limpiar filas con valores nulos (si los hubiera).

**2. Creación de variables nuevas (NumPy/Pandas):**

* Crear una nueva columna llamada consumo\_alcohol\_por\_persona = litros\_bebida\_alcoholica / n\_asistentes.
* Crear otra columna llamada accidentes\_por\_mil\_asistentes = (accidentes\_transito / n\_asistentes) \* 1000.

**3. Análisis estadístico (Pandas):**

* Calcular el promedio de **consumo\_alcohol\_por\_persona** por región.
* Calcular el promedio de accidentes\_por\_mil\_asistentes por región.
* Identificar la región con **mayor consumo de alcohol per cápita** y la región con **más accidentes por mil asistentes**.

**4. Visualización (Matplotlib):**

* Hacer un **gráfico de dispersión (scatter plot)** con:
  + Eje X consumo\_alcohol\_por\_persona.
  + Eje Y accidentes\_por\_mil\_asistentes.
  + Cada punto debe representar un **evento**.
  + **Colorear los puntos según la región** para comparar visualmente.
  + Agregar **título y etiquetas** claras.

**5. Interpretación (explicación escrita):**

* ¿Se observa una **tendencia**? (ejemplo: mientras aumenta el consumo de alcohol por persona, ¿también aumentan los accidentes?).
* ¿Qué **región se ve más crítica**?
* ¿Qué **recomendaciones** podrían hacer las autoridades en base a los datos?

Codigo

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**import** **matplotlib.pyplot** **as** **plt**

# Cargar el dataset

df = pd.read\_csv("fiestas.csv")

# Crear nuevas columnas

df['consumo\_alcohol\_por\_persona'] = df['litros\_bebida\_alcoholica'] / df['n\_asistentes']

df['accidentes\_por\_1000'] = (df['accidentes\_transito'] / df['n\_asistentes']) \* **1000**

print("**\n**")

print("Promedio de consumo por persona:", df['consumo\_alcohol\_por\_persona'].mean())

print("Promedio de accidentes por 1000 personas:", df['accidentes\_por\_1000'].mean())

print("Correlación entre consumo y accidentes:", df[['consumo\_alcohol\_por\_persona', 'accidentes\_por\_1000']].corr().iloc[**0**, **1**])

print("**\n**")

plt.figure(figsize=(**8**, **6**))

plt.scatter(df['consumo\_alcohol\_por\_persona'], df['accidentes\_por\_1000'], color='crimson', alpha=**0.7**, edgecolors='black')

plt.title("Consumo de Alcohol vs Accidentes por 1000 Personas")

plt.xlabel("Consumo de Alcohol por Persona (litros)")

plt.ylabel("Accidentes por cada 1000 Asistentes")

plt.grid(**True**)

plt.tight\_layout()

plt.show()

print("**\n**")

Texto

El contenido generado por IA puede ser incorrecto.

Se crean dos nuevas columnas: consumo\_alcohol\_por\_persona y accidentes\_por\_1000. La primera columna calcula el promedio de litros de bebida alcohólica consumidos por cada asistente. La segunda, normaliza la cantidad de accidentes de tránsito por cada 1000 asistentes. Estas métricas permiten comparar eventos de diferentes tamaños de una manera más justa.

**Visualización (Matplotlib)**

Codigo

**import** **pandas** **as** **pd**

**import** **matplotlib.pyplot** **as** **plt**

# Cargar el dataset

df = pd.read\_csv("fiestas.csv")

# columnas nuevas

df['consumo\_alcohol\_por\_persona'] = df['litros\_bebida\_alcoholica'] / df['n\_asistentes']

df['accidentes\_por\_mil\_asistentes'] = (df['accidentes\_transito'] / df['n\_asistentes']) \* **1000**

regiones = df['region'].unique()

colores = plt.cm.tab20.colors

color\_map = {region: colores[i % len(colores)] **for** i, region **in** enumerate(regiones)}

plt.figure(figsize=(**10**, **6**))

**for** region **in** regiones:

datos = df[df['region'] == region]

plt.scatter(

datos['consumo\_alcohol\_por\_persona'],

datos['accidentes\_por\_mil\_asistentes'],

color=color\_map[region],

label=region,

alpha=**0.7**,

edgecolors='black'

)

plt.title("Consumo de Alcohol vs Accidentes por 1000 Asistentes, según Región")

plt.xlabel("Consumo de Alcohol por Persona (litros)")

plt.ylabel("Accidentes por cada 1000 Asistentes")

plt.legend(title="Región", bbox\_to\_anchor=(**1.05**, **1**), loc='upper left')

plt.grid(**True**)

plt.tight\_layout()

plt.show()

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

**Análisis del gráfico: Consumo de Alcohol vs Accidentes por 1000 Asistentes, según Región**

**1. ¿Se observa una tendencia?**

Sí, se observa una tendencia moderada: a medida que aumenta el consumo de alcohol por persona, también tiende a aumentar la tasa de accidentes por cada 1000 asistentes. Aunque la mayoría de los eventos se concentran en niveles bajos de consumo y accidentes, hay varios casos excepcionales con valores elevados que refuerzan esta relación. Sin embargo, la dispersión de los puntos indica que no es una correlación perfecta, lo que sugiere que otros factores también influyen (como clima, tipo de evento, medidas de seguridad, etc.).

**2. ¿Qué región se ve más crítica?**

**La región Metropolitana destaca como una de las más críticas:**

Presenta eventos con alto consumo de alcohol por persona.

* Algunos de esos eventos también tienen tasas elevadas de accidentes por 1000 asistentes.
* Además, por su densidad poblacional y volumen de eventos, cualquier desviación tiene mayor impacto.
* También se observan puntos preocupantes en Valparaíso y Biobío, aunque con menor frecuencia.

**3. ¿Qué recomendaciones podrían hacer las autoridades?**

* Para el Ministerio de Transportes:
* Refuerzo de controles viales en regiones con alta concentración de eventos y consumo elevado.
* Monitoreo en tiempo real de eventos masivos para detectar riesgos emergentes.

**Para el Ministerio de Salud:**

* Educación preventiva sobre consumo responsable, especialmente en comunas con historial de accidentes.
* Intervenciones comunitarias en regiones con mayor riesgo, como Metropolitana y Valparaíso.