

Proyecto Integrador: Análisis Probabilístico del Clima en Seattle (2012)

Objetivo general:

Aplicar conocimientos de estadística y probabilidad para analizar el comportamiento del clima en Seattle, utilizando un enfoque basado en eventos, relaciones causales y visualizaciones. El análisis debe permitir generar recomendaciones basadas en datos para actividades al aire libre, campañas de marketing y gestión de riesgos climáticos.

Dataset base

Fuente CSV:


 <https://raw.githubusercontent.com/vega/vega-datasets/main/data/seattle-weather.csv>

```
python
CopiarEditar
import pandas as pd

# 1. Cargar y preparar el dataset
df = pd.read_csv("https://raw.githubusercontent.com/vega/vega-datasets/main/data/seattle-weather.csv")
df["date"] = pd.to_datetime(df["date"])
df["month"] = df["date"].dt.month
df["year"] = df["date"].dt.year
df["day"] = df["date"].dt.day
```

Retos Integradores

Reto 1 - Análisis de eventos climáticos extremos

 **Propósito:** Aplicar la definición de eventos y cálculo de probabilidades simples y condicionales.

Instrucciones:

- Define los eventos:
 - A: "precipitation ≥ 1 "
 - B: "weather == 'rain'"
 - C: "temp_max ≥ 25 "
 - Calcula las probabilidades:
 - $P(A)P(A)P(A)$, $P(B)P(B)P(B)$, $P(C)P(C)P(C)$
 - $P(A|B)P(A \mid B)P(A|B)$, $P(A|C)P(A \mid C)P(A|C)$
 - Grafica:
 - Un **gráfico de barras** con la frecuencia mensual de días con precipitación ≥ 1 .
 - Un **heatmap** de $P(A|BAC)P(A \mid B \land C)P(A|BAC)$ por mes.
-


Reto 2 - Estacionalidad y comportamiento del clima

 **Propósito:** Estudiar cómo varían las condiciones climáticas a lo largo del año.

Instrucciones:

1. Calcula:
 - o La proporción de días `"weather = 'sun'"` para cada mes del año.
 - o El promedio mensual de `precipitation` y `temp_max`.
 2. Grafica:
 - o **3 líneas** (una para cada variable: `sun_days`, `temp_max`, `precip`) a lo largo de los 12 meses.
 3. Interpreta:
 - o ¿Qué meses son más adecuados para eventos al aire libre?
-


Reto 3 - Muestreo y sesgo

 **Propósito:** Comparar una muestra aleatoria con una muestra sesgada para observar efectos en las conclusiones.

Instrucciones:

1. Toma dos muestras de tamaño 60:
 - o Aleatoria simple (`sample(n=60)`)
 - o Sesgada: solo días con `"temp_max ≥ 25"`
 2. Calcula para ambas:
 - o Media y varianza de `precipitation`
 - o Porcentaje de días sin lluvia (`precipitation == 0`)
 3. Grafica:
 - o **Boxplot** comparativo entre ambas muestras.
 4. Reflexiona:
 - o ¿Qué diferencia hay entre las muestras? ¿Qué riesgo implica un sesgo en la toma de decisiones?
-


Reto 4 - Comparación entre clima real e ideal

 **Propósito:** Evaluar qué tan cerca está el clima real de un escenario ideal para trabajar al aire libre.

Instrucciones:

1. Crea una columna `"ideal_precip"` con valor 0.0 (sin lluvia).
 2. Calcula:
 - o Media, varianza y desviación estándar de `precipitation` y `ideal_precip`
 3. Grafica:
 - o Un **histograma** y un **kdeplot** de ambas columnas.
 4. Interpreta:
 - o ¿Cuán diferente es el clima real del ideal en Seattle?
-


Reto 5 - Regresión: ¿la temperatura predice precipitación?

 **Propósito:** Aplicar un modelo de regresión para estimar la relación entre variables meteorológicas.

Instrucciones:

1. Ajusta un modelo de regresión lineal:
`precipitation ~ temp_max`
 2. Reporta:
 - o Coeficiente de la pendiente
 - o Valor de R^2
 - o Significancia estadística
 3. Grafica:
 - o Un **gráfico de dispersión** + línea de regresión (`sns.lmplot`)
 - o Opcional: residuos
 4. Evalúa:
 - o ¿Qué tan útil es este modelo para predecir si lloverá?
-

Reto 6 - Construcción de un árbol de probabilidad

 **Propósito:** Descomponer eventos encadenados con probabilidad condicional.

Instrucciones:

1. Usa los siguientes niveles:
 - o Nivel 1: `weather ∈ {sun, rain, fog}`
 - o Nivel 2: `temp_max ≥ 25`
 - o Nivel 3: `precipitation ≥ 1`
 2. Calcula:
 - o Probabilidades conjuntas y condicionales:
Ej. $P(\text{precip} \geq 1 | \text{weather} = \text{'rain'} \wedge \text{temp} \geq 25) = \frac{P(\text{precip} \geq 1 \wedge \text{weather} = \text{'rain'} \wedge \text{temp} \geq 25)}{P(\text{weather} = \text{'rain'} \wedge \text{temp} \geq 25)}$
 3. Crea:
 - o Un **diagrama de árbol** textual con ramas o una tabla que refleje los caminos.
 4. Reflexiona:
 - o ¿Qué combinación de condiciones representa mayor riesgo climático?
-

Herramientas sugeridas

Herramienta	Usos recomendados
<code>pandas</code>	Filtrado, agrupamiento, cálculos de proporción
<code>seaborn</code> , <code>matplotlib</code>	Gráficos: barras, líneas, regresión, KDE
<code>statsmodels.formula.api</code>	Ajuste de modelos lineales (<code>ols</code>)
<code>numpy</code>	Cálculos adicionales (varianza, media)
