



Practica 19 - Pandas KPI's

📅 Date	@November 19, 2025
✓ check	<input type="checkbox"/>

Parte 1 - KPIs con Pandas

🎯 Contexto del Negocio

Trabajas como Data Value Engineering en "**LearnHub**", una plataforma de cursos online. Tu tarea es analizar el comportamiento de los usuarios y calcular KPIs clave usando pandas para la toma de decisiones estratégicas.

📁 Dataset Inicial

```
import pandas as pd
import numpy as np

# Configurar semilla para reproducibilidad
```

```
np.random.seed(42)

# Generar 100 registros
n_registros = 100

# Generar datos
data = {
    'usuario_id': range(1, n_registros + 1),

    'fecha_registro': pd.date_range('2023-01-01', periods=n_registros, freq='3D'),

    'fecha_primera_compra': [
        pd.Timestamp('2023-01-01') + pd.Timedelta(days=np.random.randint(1, 30))
        if np.random.random() > 0.25 else None
        for _ in range(n_registros)
    ],
    'fecha_ultima_actividad': pd.date_range('2024-10-01', periods=n_registros,
freq='2D'),
    'plan': np.random.choice(
        ['Free', 'Basic', 'Standard', 'Premium'],
        n_registros,
        p=[0.25, 0.30, 0.25, 0.20]
    ),
    'precio_plan': None, # Se calculará después según el plan
    'canal_adquisicion': np.random.choice(
        ['Google Ads', 'Facebook Ads', 'Instagram Ads', 'Organic', 'Referral', 'Email Marketing'],
        n_registros,
        p=[0.20, 0.15, 0.15, 0.25, 0.15, 0.10]
    ),
}
```

```

'gasto_marketing': None, # Se calculará después según el canal

'cursos_completados': np.random.randint(0, 15, n_registros),

'total_cursos_inscritos': None, # Se calculará después

'tiempo_plataforma_horas': np.random.uniform(0, 200, n_registros).round
(2),


'pais': np.random.choice(
    ['España', 'México', 'Argentina', 'Colombia', 'Chile', 'Perú', 'Estados Unidos'],
    n_registros,
    p=[0.20, 0.20, 0.15, 0.15, 0.10, 0.10, 0.10]
),


'estado_suscripcion': np.random.choice(
    ['Activo', 'Cancelado', 'Pausado'],
    n_registros,
    p=[0.65, 0.25, 0.10]
),


'numero_recomendaciones': np.random.randint(0, 8, n_registros),


'certificados_obtenidos': None # Se calculará después
}

df = pd.DataFrame(data)

# Asignar precios según el plan
precios = {'Free': 0, 'Basic': 29.99, 'Standard': 59.99, 'Premium': 99.99}
df['precio_plan'] = df['plan'].map(precios)

# Asignar gasto de marketing según el canal (solo para canales pagos)
gastos_canal = {

```

```

'Google Ads': lambda: np.random.uniform(10, 25),
'Facebook Ads': lambda: np.random.uniform(8, 20),
'Instagram Ads': lambda: np.random.uniform(7, 18),
'Organic': lambda: 0,
'Referral': lambda: np.random.uniform(0, 5),
'Email Marketing': lambda: np.random.uniform(2, 8)
}

df['gasto_marketing'] = df['canal_adquisicion'].apply(
    lambda x: round(gastos_canal[x](), 2)
)

# Cursos inscritos siempre >= cursos completados
df['total_cursos_inscritos'] = df['cursos_completados'] + np.random.randint(0,
10, n_registros)

# Certificados obtenidos (solo para cursos completados, no todos los cursos
dan certificado)
df['certificados_obtenidos'] = df['cursos_completados'].apply(
    lambda x: np.random.randint(0, max(1, x + 1)) if x > 0 else 0
)

# Los usuarios Free no tienen fecha de compra ni pagan
df.loc[df['plan'] == 'Free', 'fecha_primera_compra'] = None

# Ajustar fechas de última actividad para usuarios cancelados (más antiguas)
mask_cancelados = df['estado_suscripcion'] == 'Cancelado'
df.loc[mask_cancelados, 'fecha_ultima_actividad'] = pd.date_range(
    '2024-06-01',
    periods=mask_cancelados.sum(),
    freq='3D'
)

print("Dataset creado exitosamente!")
print(f"Total de registros: {len(df)}")
print(f"\nPrimeros registros:")

```

```
print(df.head())
print(f"\nInformación del dataset:")
print(df.info())
```

Ejercicios

Ejercicio 1: Tasa de Conversión Global

Objetivo: Calcular el porcentaje de usuarios registrados que se convirtieron de plan gratuito a plan de pago.

Ecuación:

$$\text{Tasa de Conversión (\%)} = (\text{Usuarios con Plan de Pago} / \text{Total de Usuarios Registrados}) \times 100$$

Significado teórico:

Este KPI mide la efectividad del modelo freemium de la plataforma. Una tasa de conversión alta indica que los usuarios perciben suficiente valor en la versión gratuita como para justificar el pago por funcionalidades premium. Típicamente, las plataformas freemium exitosas tienen tasas de conversión entre 2-5%.

Ejercicio 2: Tiempo Promedio de Conversión

Objetivo: Para los usuarios que compraron un plan de pago, calcular el tiempo promedio (en días) que transcurrió entre su registro y su primera compra.

Ecuación:

$$\text{Tiempo Promedio de Conversión (días)} = \text{Promedio}(\text{fecha_primera_compra} - \text{fecha_registro})$$

Nota: Calcular solo para usuarios con fecha_primera_compra no nula

Significado teórico:

Mide la velocidad del ciclo de conversión y la efectividad del funnel de ventas. Un tiempo menor indica que la propuesta de valor es clara y convincente. Este KPI

ayuda a optimizar estrategias de nurturing y permite establecer el momento óptimo para campañas de conversión.

Ejercicio 3: ARPU (Average Revenue Per User)

Objetivo: Calcular el ingreso promedio por usuario considerando únicamente usuarios que tienen un plan de pago.

Ecuación:

$$\text{ARPU} = \frac{\text{Suma Total de Precios de Planes}}{\text{Número de Usuarios con Plan de Pago}}$$

Significado teórico:

Indica el valor monetario promedio que genera cada cliente de pago. Es fundamental para proyecciones financieras, evaluación de estrategias de pricing y comparación con el CAC. Un ARPU creciente puede indicar éxito en estrategias de upselling.

Ejercicio 4: CAC por Canal de Adquisición

Objetivo: Calcular el Costo de Adquisición de Cliente (CAC) para cada canal de marketing, considerando únicamente los usuarios que se convirtieron a planes de pago.

Ecuación:

$$\text{CAC por Canal} = \frac{\text{(Gasto Total de Marketing en el Canal)}}{\text{(Número de Usuarios de Pago Adquiridos por ese Canal)}}$$

Significado teórico:

Identifica la eficiencia de cada canal para adquirir clientes de pago (no solo registros). Permite optimizar el presupuesto de marketing invirtiendo más en canales con CAC bajo. Un CAC debe compararse siempre con el LTV para determinar la rentabilidad del canal.

Ejercicio 5: Distribución de Ingresos por Plan

Objetivo: Calcular qué porcentaje de los ingresos totales representa cada tipo de plan (Basic, Standard, Premium).

Ecuación:

$$\text{Participación del Plan (\%)} = (\text{Ingresos del Plan Específico} / \text{Ingresos Totales}) \times 100$$

donde:

$$\text{Ingresos del Plan} = \text{Cantidad de Usuarios del Plan} \times \text{Precio del Plan}$$

Significado teórico:

Muestra la distribución de ingresos entre los diferentes tiers de producto. Ayuda a identificar qué plan es el motor de ingresos y cuál podría necesitar optimización. Esta información es crucial para decisiones sobre desarrollo de producto y estrategias de pricing.

Ejercicio 6: Tasa de Finalización de Cursos (Completion Rate)

Objetivo: Calcular el porcentaje promedio de cursos que los usuarios completan respecto a los cursos en los que se inscriben.

Ecuación:

$$\text{Completion Rate (\%)} = \text{Promedio}(\text{cursos_completados} / \text{total_cursos_inscritos}) \times 100$$

Nota: Excluir usuarios con 0 cursos inscritos para evitar divisiones por cero

Significado teórico:

Mide el nivel de engagement y la calidad percibida del contenido. Una tasa alta indica que los usuarios encuentran valor en los cursos y se mantienen comprometidos. Una tasa baja puede señalar problemas de calidad del contenido, dificultad inadecuada o falta de motivación.

Ejercicio 7: ROI por Canal de Marketing

Objetivo: Calcular el Retorno de Inversión para cada canal de adquisición considerando los ingresos generados vs. el gasto en marketing.

Ecuación:

$$\begin{aligned}\text{Ingresos por Canal} &= \text{Suma}(\text{precio_plan}) \text{ de usuarios adquiridos por ese canal} \\ \text{Gasto por Canal} &= \text{Suma}(\text{gasto_marketing}) \text{ del canal} \\ \text{ROI por Canal (\%)} &= ((\text{Ingresos por Canal} - \text{Gasto por Canal}) / \text{Gasto por Canal}) \times 100\end{aligned}$$

Significado teórico:

Determina la rentabilidad real de cada canal de marketing. Un ROI positivo indica que el canal es rentable; un ROI superior al 200% generalmente se considera excelente en marketing digital. Este KPI es crítico para la asignación eficiente del presupuesto de marketing.

Ejercicio 8: Tasa de Retención (Usuarios Activos)

Objetivo: Calcular el porcentaje de usuarios con plan de pago que mantienen su suscripción activa.

Ecuación:

$$\text{Tasa de Retención (\%)} = (\text{Usuarios con Estado "Activo" y Plan de Pago} / \text{Total de Usuarios con Plan de Pago}) \times 100$$

Significado teórico:

Mide la capacidad de la plataforma para mantener a sus clientes de pago. Una alta tasa de retención indica satisfacción del cliente y un producto-mercado fit sólido. Es más económico retener clientes que adquirir nuevos, por lo que este KPI es fundamental para la sostenibilidad del negocio.

Ejercicio 9: Engagement Score por País

Objetivo: Crear un score de engagement por país combinando múltiples métricas: cursos completados, tiempo en plataforma y certificados obtenidos.

Ecuación:

Engagement Score = Promedio por País de:
(cursos_completados × 0.4) +
(tiempo_plataforma_horas / 10 × 0.3) +
(certificados_obtenidos × 0.3)

Significado teórico:

Un score compuesto permite evaluar el nivel de compromiso de usuarios en diferentes mercados geográficos. Los pesos asignados reflejan la importancia relativa de cada métrica (completar cursos es el indicador más fuerte de engagement). Este KPI ayuda a identificar mercados con mejor product-market fit y a personalizar estrategias por región.

Ejercicio 10: Viral Coefficient (K-Factor)

Objetivo: Calcular el coeficiente viral estimado basado en el número de recomendaciones realizadas y la tasa de conversión de referidos.

Ecuación:

Promedio de Invitaciones por Usuario = Promedio(numero_recomendaciones)

Tasa de Conversión de Referidos = Usuarios del Canal "Referral" con Plan de Pago / Total de Usuarios del Canal "Referral"

K-Factor = Promedio de Invitaciones × Tasa de Conversión de Referidos

Significado teórico:

El K-Factor mide el potencial de crecimiento viral de la plataforma. Un $K > 1$ indica crecimiento exponencial orgánico (cada usuario trae más de un usuario nuevo). Un K entre 0.5-1 es muy saludable y reduce significativamente el CAC. Este KPI es fundamental para plataformas que dependen del crecimiento orgánico y el marketing de referidos.



Instrucciones Generales

1. Conversión de tipos de datos:

- Asegúrate de que `fecha_registro`, `fecha_primera_compra` y `fecha_ultima_actividad` sean tipo `datetime`

2. Manejo de valores nulos:

- Algunos usuarios no tienen `fecha_primera_compra` (usuarios Free o que no compraron)
- Usa `.dropna()` o `.notna()` cuando sea necesario
- Documenta cómo manejas los nulos en cada ejercicio

3. Buenas prácticas:

- Crea variables intermedias con nombres descriptivos
- Documenta cada paso con comentarios
- Utiliza `groupby()`, `agg()`, `merge()` cuando sea apropiado
- Redondea los resultados a 2 decimales para legibilidad

4. Validación:

- Verifica que los resultados tengan sentido lógico
- Comprueba que no hay divisiones por cero
- Valida que los porcentajes estén entre 0 y 100

5. Presentación:

- Muestra los resultados con formato claro
- Incluye unidades (%\$, \$, días, etc.)
- Añade interpretaciones breves de los resultados

Parte 2 - Análisis descriptivo de PokemonGo + PowerBI Report

🎯 Contexto del Negocio

Trabajas como Data Value Engineering en una empresa de gaming. Tu tarea es realizar un análisis descriptivo, encontrar valor en los datos y calcular KPI's clave usando pandas para la toma de decisiones estratégicas.

Dataset Inicial

Descarga el siguiente fichero:

<https://tajamar365.sharepoint.com/:u/s/3431-MasterIA2025-2026/EZltFwpIY8VLlaUJNqTup5oBev0IE02rBh3CCQrB-xEaQw?e=6xv4pi>

Instrucciones.

Una vez descargado el fichero y cargado en tu notebook:

- 2.1 Efectuar preprocesamiento y limpieza de datos
- 2.2 Realizar un análisis descriptivo. Explicar en qué consiste la dataset, qué significa cada columna, realizar estadísticas básicas (ver clase 32), identifica outliers.
- 2.3 Identificar Insights o datos de valor.
- 2.4 Construye KPIs
- 2.5 Aplicar Feature Engineering "A priori", es decir: ¿Existirán columnas irrelevantes desde el punto de vista descriptivo o diagnóstico?. ¿Es posible combinar varias columnas en una única?
- 2.6 Elaborar un reporte en PowerBI que muestre: Insights (algunos gráficos que demuestren datos importantes) , KPI's, estadísticas descriptivas básicas.
- 2.7 Elaborar una presentación en PowerPoint destacando los aspectos más relevantes del estudio realizado. Imagina que vas a realizar la presentación ante el equipo de Analistas de Negocios (BI Analyst), Analistas de datos y directivos "Level C" no técnicos.