# **Prueba de Desempeño – Analítica de Datos (RIWI Sport)**

## **Caso de uso**

RIWI Sport comercializa artículos deportivos en Colombia. La Dirección de Analítica requiere medir el comportamiento de clientes y ventas para orientar acciones comerciales por categoría y ciudad, y tomar decisiones basadas en datos.

## **Objetivo**

Bajo tu **rol como Analista de Datos**, deberás construir un análisis reproducible en Jupyter Notebook que:

* Se conecte a una base de datos PostgreSQL (cargada desde el archivo RiwiSport.sql).
* Consulte tablas relevantes (customer, order, order\_item, product, category).
* Aplique Pandas y NumPy para calcular KPIs (tendencia central y dispersión) y KPIs de negocio.
* Genere visualizaciones con Matplotlib y Seaborn.
* Redacte al menos un insight accionable con una recomendación breve para RIWI Sport.

## **Funcionalidades principales**

* Conexión a PostgreSQL desde Python (**sqlalchemy** o **psycopg2**).
* Consultas SQL a las tablas: customer, order, order\_item, product, category.
* Extracción a **DataFrames** y **limpieza mínima** (tipos, nulos, duplicados).
* Cálculo de **KPIs** estadísticos y de negocio (ver Requisitos).
* Visualizaciones en **Matplotlib** y **Seaborn** que respalden los hallazgos.
* Redacción de un **informe ejecutivo** con al menos un insight y su recomendación.

## **Requisitos**

### **1) Conexión y datos**

* Restaurar el archivo RiwiSport.sql en un servidor local de **PostgreSQL**.
* Establecer conexión desde Python (sqlalchemy o psycopg2).
* Verificar que las **tablas y seeds** se hayan cargado correctamente.

### **2) Exploración y consultas**

* Consultar tablas relevantes: customer, order, order\_item, product, category.
* Construir una **tabla de ventas** a nivel de ítem y por pedido, integrando producto y categoría.
* Obtener agregados por **ciudad**, **categoría**, **producto** y **cliente** cuando aplique.

### **3) Análisis con Pandas/NumPy (KPIs obligatorios)**

* **Tendencia central:** media, mediana, moda del gasto (por pedido/cliente).
* **Dispersión:** varianza, desviación estándar; opcional: IQR (rango intercuartílico).
* **KPIs de negocio:**
  + **Ticket Promedio (AOV)** por pedido y por cliente.
  + **Top 5 categorías** por ventas y **Top 5 productos** por cantidad/ingreso.
  + **Producto con mayor variabilidad de precio** (si los datos lo permiten).

### **4) Visualizaciones (Matplotlib y Seaborn)**

* **Histograma** del gasto por cliente.
* **Boxplot** de ventas por categoría (para observar distribución y outliers).
* **Barras** para Top 5 categorías/productos por ventas.

### **5) Insight y storytelling**

* Redactar **al menos 1 insight no trivial** basado en la evidencia y una **recomendación breve** (qué harías y por qué).
  + Ejemplo: “Running muestra mayor dispersión de precios; segmentar oferta en gama básica vs. premium en ciudades con alta variabilidad.”

## **Criterios de aceptación**

* La base de datos se **restaura** y la **conexión** desde Python funciona.
* Las **consultas SQL** integran correctamente las tablas y responden al caso de negocio.
* Los **KPIs** estadísticos y de negocio están **bien calculados** y **explicados**.
* Las **visualizaciones** son pertinentes, legibles y apoyan los hallazgos.
* El **insight** es claro, basado en datos y útil para la decisión.
* El **notebook** es reproducible (celdas ordenadas, dependencias claras) y el **informe** está bien estructurado.

## **Entrega y documentación**

* **Notebook**: analisis\_RIWI\_Sport\_Nombre-Apellido.ipynb con conexión, SQL, Pandas/NumPy, KPIs, gráficos e insight.
* **Informe**: informe\_RIWI\_Sport.pdf o .md con **resumen ejecutivo**, **KPIs** y **visualizaciones** clave.
* (Opcional) Ejemplo de **.env** para la conexión a BD.
* **README** con instrucciones para restaurar la BD, configurar variables y ejecutar el notebook. Con dependencias mínimas (sugeridas): pandas, numpy, sqlalchemy, psycopg2-binary, matplotlib, seaborn, python-dotenv.

## **Consideraciones generales**

* Mantener buenas prácticas: nombres claros, orden de celdas, comentarios donde agreguen valor.
* Si el dataset/BD no permite calcular exactamente algún KPI, justificar y proponer una aproximación válida.

## **Puntos Extras (opcionales)**

* Notebook parametrizable (variables de entorno, funciones reutilizables).
* Métricas adicionales: **IQR**, **coeficiente de variación**, **pruebas de normalidad**.
* **Segmentación** exploratoria por ciudad/categoría con conclusiones por segmento.

## **Recursos**

* Documentación oficial de **PostgreSQL** y **SQL**.
* Documentación oficial de **Python**, **Pandas**, **NumPy**.
* Documentación oficial de **Matplotlib** y **Seaborn**.
* Archivo **estructuraypoblado.sql** (DDL + seeds) para cargar la base de datos.