🍺 Bar Intelligence: Data-Driven Optimization for "'CONFIDENCIAL' Bar"

| STATUS | **COMPLETED** | PYTHON | **3.11** | STACK | **PANDAS \| SEABORN \| SCIKITLEARN** |

📖 Executive Summary

This project delivers a comprehensive **Business Intelligence & Data Science analysis** for "'CONFIDENCIAL' Bar", based on **8,800+ transactional records** (June-Nov 2025).

The objective was to transform raw sales logs into actionable strategies. By applying **Data Cleaning**, **Statistical Analysis**, and **Machine Learning (K-Means Clustering)**, we identified opportunities to **optimize inventory by 80%**, adjust pricing strategies based on **demand elasticity**, and improve staff scheduling efficiency.

📊 Key Strategic Insights & Visualizations

1. Inventory Optimization (Pareto Principle)

The Insight: The business follows a strict Pareto distribution. A single product (*Club 850ml*) drives the majority of the revenue. The top 6 products generate nearly **50% of total income**.

- **Action:** "Ring-fence" the stock for these critical items. Zero tolerance for stockouts on the top 6 SKUs.

Pareto Analysis *(Figure 1: Top 20 Products driving 80% of cumulative revenue)*

2. Customer Segmentation (K-Means Clustering)

The Insight: Using Unsupervised Learning, we identified three distinct customer profiles with unique temporal behaviors:

- **Light Consumers:** High volume, low ticket. Scattered arrival times (After-office).

- **Standard Consumers:** The core profitability group. Concentrated during "Prime Time" (20:00-23:00).

- **VIP / Whales:** Outlier spenders. They arrive late (**22:00+**).

- **Action:** Assign senior staff to late shifts to cater to VIPs and run aggressive "Happy Hours" (16:00-19:00) to consolidate the Light segment.

Customer Segmentation *(Figure 2: K-Means Clusters based on Purchase Volume vs. Total Spend)*

3. Price Elasticity & Sensitivity

The Insight: Demand sensitivity varies significantly by product category.

- **Inelastic (Premium):** Products like *Fraile* and *Zhumir Pink* maintain sales even at higher prices. **Opportunity:** Margin increase (5-10%).

- **Elastic (Economy):** Products like *RTD* show sharp sales drops when prices rise. **Opportunity:** Use as promotional hooks only.

Price Elasticity *(Figure 3: Regression analysis showing demand sensitivity to price changes)*

4. Operational Efficiency (Hourly Demand)

The Insight: Heatmaps revealed clear "dead zones" and "peak zones".

- **Action:** Staffing schedules were optimized to reduce labor costs during the 14:00-17:00 window and maximize coverage during the 21:00-00:00 peak.

Hourly Demand *(Figure 4: Revenue Curve and Order Density Heatmap)*

5. External Factors (Weather & Events)

The Insight:

- **Events:** National Soccer Matches ("La Tri") generate a measurable statistical lift in sales vs. benchmark days.
- **Weather:** While general sales are stable, specific categories (Beer vs. Spirits) show correlation with temperature shifts.

Weather Impact *(Figure 5: Correlation between Daily Temperature and Revenue)*

🛠 Methodology & Tech Stack

The project followed a structured Data Science pipeline:

1. **Data Sanitization (ETL):**
   - Handling missing values (Imputation).
   - Type casting (`datetime`, `categorical`).
   - Normalization of product names (String manipulation).
2. **Feature Engineering:**
   - **Enrichment:** Integrated **Weather Data** (Open-Meteo API) and **Holidays** (Python `holidays` library).
   - **Temporal Features:** Extraction of Day, Hour, Weekend flags.
3. **Exploratory Data Analysis (EDA):**
   - Weekly/Hourly seasonality analysis.
   - Market Basket Analysis (Co-occurrence matrix).
4. **Modeling:**
   - **K-Means Clustering:** For customer segmentation (`sklearn`).
   - **Linear Regression:** For price elasticity calculation (`scipy.stats`).

📁 Repository Structure

```
DS_BI_BAR/
|
```

```
├── notebooks/
│   ├── doc/                          # Generated plots and report assets
│       ├── customer_segmentation_clusters.png
│       ├── holiday_impact_analysis.png
│       ├── hourly_demand_summary.png
│       ├── market_basket_heatmap.png
│       ├── pareto_analysis.png
│       ├── price_elasticity_grid.png
│       ├── soccer_match_impact.png
│       ├── weather_impact_by_category.png
│       ├── weather_impact_general.png
│       └── weekly_analysis_summary.png
│   └── DS_BI_BAR.ipynb               # Main Jupyter Notebook (Source Code)
│
├── ventas/                          # Raw Data Source (CSV)
│   └── ControlBotBar - EntradaDiaria.csv
│
├── .gitignore                       # Git ignore rules
├── requirements
└── README.md
```